

清华理解科学丛书

(套装共10册)



清华大学出版社

总 目 录

[因为星星在那里：科学殿堂的砖与瓦](#)

[物含妙理总堪寻：从爱因斯坦到霍金](#)

[小楼与大师：科学殿堂的人和事](#)

[从奇点到虫洞：广义相对论专题选讲](#)

[上下百亿年：太阳的故事](#)

[日出：量子力学与相对论](#)

[那颗星星不在星图上：寻找太阳系的疆界](#)

[看不见的星：黑洞与时间之河](#)

[运动的旋律与变化的世界（大众科普著作） / 理解科学丛书](#)

[迷蒙星空：探天之路](#)

理解科学丛书·卢昌海科普著作

BECAUSE
STARS
Bricks and Tiles
of the Temple of Science ARE THERE

因为星星在那里

科学殿堂的砖与瓦

卢昌海◎著

数学的纯粹、物理的绚烂……
原子里的奥秘、星空中的未知……
迷人的科学知识、锐利的科学方法……

清华大学出版社

理解科学丛书

**Because Stars are There:
Bricks and Tiles of the Temple of Science**

因为星星在那里：科学殿堂的砖与瓦

卢昌海 著

清华大学出版社
北 京

作者简介

卢昌海，出生于杭州，本科就读于复旦大学物理系，毕业后赴美留学，于2000年获美国哥伦比亚大学物理学博士学位，目前旅居纽约。著有《那颗星星不在星图上：寻找太阳系的疆界》、《太阳的故事》、《黎曼猜想漫谈》（获第七届吴大猷科学普及著作原创类金签奖）、《从奇点到虫洞：广义相对论专题选讲》、《小楼与大师：科学殿堂的人和事》（入选“2014中国好书”），并曾在《南方周末》、《科学画报》、《现代物理知识》、《数学文化》（任特约撰稿人）等报纸、杂志上发表一百多篇科普及专业科普作品。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

因为星星在那里：科学殿堂的砖与瓦 / 卢昌海著. ——北京：清华大学出版社，2015（理解科学丛书）

ISBN 978-7-302-40066-0

I. ①因... II. ①卢... III. ①科学知识—青少年读物 IV. ①Z228.1
②N49

中国版本图书馆CIP数据核字（2015）第089273号

责任编辑：邹开颜

封面设计：

责任校对：刘玉霞

责任印制：

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者：

装 订 者：

经 销：全国新华书店

开 本：165mm×240mm

印 张：17

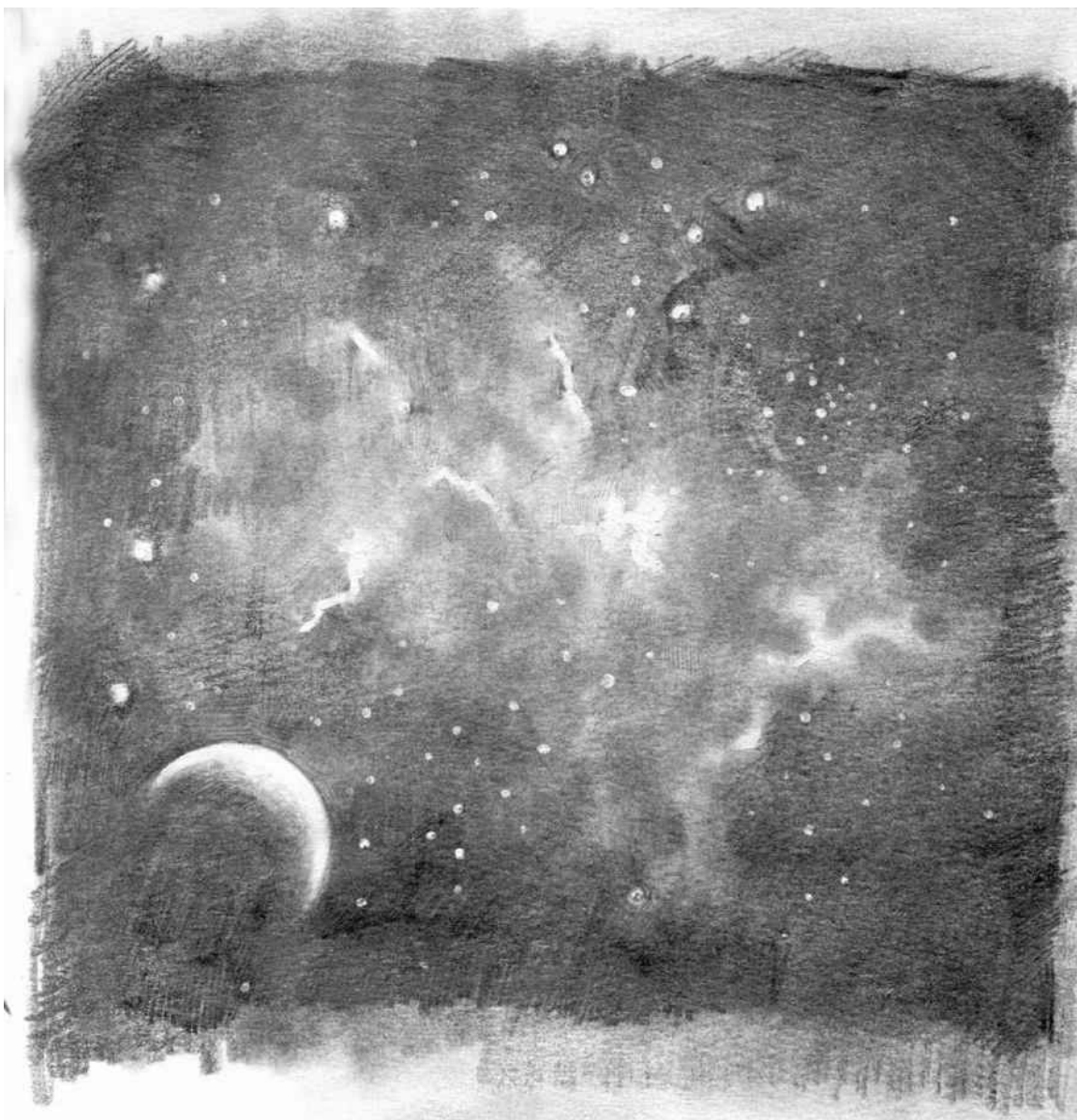
字 数：242千字

版 次：2015年6月第1版

印 次：2015年6月第1次印刷

印 数：1～ 000

产品编号：057284-01



绘画：张京

谨以本书献给我的家人

序言

作为一位出版了四本书的作者，如果要用一句话来概括写书的感觉的话，那就是：写书比写文章累。这貌似是一句显而易见的大白话，对我这种在写作上有一定兴趣，甚至以写作为乐的人来说，却是一种只有经历过了才意识到的新感觉。这新感觉的起因也是一句显而易见的大白话，那就是：书比文章长。不过，这个“长”对我来说与其说是篇幅之长，不如说是指所费时间之长。因为在一本书的写作过程中，我得不断约束自己的阅读兴趣，把主要精力投注于单一主题。另一方面，我的写作速度又比较慢（或美其名曰“认真”），从而使得写作过程往往长到了对题材的兴趣将尽而书稿远未完成的程度。这时候，写书就变成了对恒心和毅力的考验，而我——很遗憾地——曾两度在这种考验面前失败过，致使《黎曼猜想漫谈》和《从奇点到虫洞》“烂尾”多年（对这一“丢人”事迹感兴趣的读者可参阅那两本书的后记），其“累”亦由此可见。

在这种感觉下，若有谁愿把我的文章汇集成书出版，让我既免除写书之累，又可得出书之乐，那对我来说简直就是“天上掉馅儿饼”的美事，几乎要让我生出一种“偷懒”的愧疚了。最近，这样的美事居然落在了我的头上——清华大学出版社愿意出版我的两本文章合集，一本收录科学史方面的文章，一本收录科普方面的文章。

兴奋之下，我很快选好了篇目，但问题来了：一堆文章汇集在一

起，以什么作为书名呢？当然，假如我是著名作者，这根本就不是问题，大可取名为“卢昌海科学史作品集”和“卢昌海科普作品集”。但对于明显不著名的我来说，就算不怕僭越地将自己的名字厚颜纳入书名，也只会成为“票房毒药”，因此必须另谋思路。读者可能会笑话我这么小的事情都不能轻松搞定，其实非独我如此，像阿西莫夫（Isaac Asimov）那样的大牌作家也常常为书名发愁呢，以至于在文章合集*The Sun Shines Bright*的简介中感慨地说，他几乎想用数字编号来作书名了——当然，他发愁的原因跟我是不同的，他那是因为作品实在太多，显而易见的书名几乎用遍了。

经过思考，为了让两本书略显对仗，我提议将科学史合集取名为《科学殿堂的人和事》，将科普合集取名为《科学殿堂的砖与瓦》。但编辑看了之后觉得这两个标题太平淡。于是我又绞尽脑汁想了半天，却没再想出什么点子来。无奈之下，我决定效仿阿西莫夫，他虽然也为书名发愁，点子可比我多多了，在*The Sun Shines Bright*的简介中做完了用数字编号作为书名的“白日梦”后，随即采用了一个颇有些取巧的办法，那就是从所汇集的文章中选取一篇的标题作为书名。现在您所看到的这两篇文章合集的书名——《小楼与大师：科学殿堂的人和事》和《因为星星在那里：科学殿堂的砖与瓦》——便也是如此而来。

关注我文章的读者或许注意到了，收录在这两本书中的某些文章是曾经在杂志或报纸上发表过的。不过，杂志和报纸大都有自己固定的风格，有时不免需要作者“削足适履”来契合之。因此，发表在杂志和报纸上的版本与我自己的版本相比大都存在一定的缺陷，比如经过编辑的改动，以及因字数所限作过删节等。此外，发表在杂志上的版本大都略去了注释及对人名和术语的英文标注等，这其中后者——即英文标注——或许并不重要，但前者——即注释——其实是颇为重要的，往往起着补

充正文、澄清歧义等诸多作用。所有这些缺陷在此次汇集成书时都尽可能予以消除了。

与以前的四本书一样，这两本书也是非常接近原稿风格的，在个别细节上甚至有可能略胜于原稿，因为编辑订正的个别错别字由于未曾标注，我未必能在阅读校样时一一察觉并在自己的版本上做出相应的订正。在尊重原稿这个最至关重要的特点上，我要再次对清华大学出版社表示感谢，感谢其对我作品及写作风格的长期——从出版第一本书至今已五年了，够得上用这个词了吧——信任和支持。

最后，希望读者们喜欢这两本新书。

目 录

[序言](#)

[第一部分 数学](#)

[孪生素数猜想](#)

[魔方与“上帝之数”](#)

[一、风靡世界的玩具](#)

[二、魔方与“上帝之数”](#)

[三、寻找“上帝之数”](#)

[ABC猜想浅说](#)

[一、什么是ABC猜想？](#)

[二、ABC猜想为什么重要？](#)

[三、ABC猜想被证明了吗？](#)

[谷歌背后的数学](#)

[一、引言](#)

[二、基本思路](#)

[三、问题及解决](#)

[四、结语](#)

[第二部分 物理](#)

[从巴西的蝴蝶到得克萨斯的飓风](#)

[一、决定论](#)

[二、早期研究](#)

[三、模拟天气](#)

[四、奇怪的结果](#)

[五、从蝴蝶到飓风](#)

[关于时钟佯谬](#)

[一、时钟佯谬简史](#)

[二、时钟佯谬简析](#)

[三、关于理想时钟](#)

[从等效原理到爱因斯坦-嘉当理论](#)

[一、等效原理](#)

[二、爱因斯坦-嘉当理论](#)

[黑洞略谈](#)

[反物质浅谈](#)

[一、一个令人苦恼的结果](#)

[二、错误描述中的正确结论](#)

[三、走错方向的电子还是走对方向的正电子？](#)

[四、从反粒子到反物质](#)

[五、宇宙的主人和客人](#)

[六、恼人的不对称之谜](#)

[七、结语](#)

[从伽利略船舱到光子马拉松](#)

[一、从相对性原理到相对论](#)

[二、破坏相对论的思路与后果](#)

[三、光子的马拉松——破坏相对论的证据？](#)

[质量的起源](#)

[一、引言](#)

[二、宇宙物质的组成](#)

[三、从机械观到电磁观](#)

[四、经典电子论](#)

[五、量子电动力学](#)

[六、质量电磁起源的破灭](#)

[七、对称性自发破缺](#)

[八、从希格斯机制到电弱统一理论](#)

[九、量子色动力学](#)

[十、同位旋与手征对称性](#)

[十一、手征对称性自发破缺](#)

[十二、贗戈德斯通粒子的质量](#)

[十三、一个93分的答案](#)

[纤维里的光和电路中的影](#)

[一、光纤，信息时代的大动脉](#)

[二、CCD，数码摄影的电子眼](#)

[石墨烯——从象牙塔到未来世界](#)

[一、来自象牙塔的新材料](#)

[二、通往未来世界的金桥](#)

[囚禁的量子，开放的应用](#)

[一、小有小的麻烦](#)

[二、囚禁的量子](#)

[三、开放的应用](#)

[第三部分 星际旅行漫谈](#)

[因为星星在那里](#)

[火箭：宇航时代的开拓者](#)

[一、引言](#)

[二、宇宙速度](#)

[三、齐奥尔科夫斯基公式](#)

[四、接近光速](#)

[五、飞向深空](#)

[生命传输机](#)

[虫洞：遥远的天梯](#)

[一、引言](#)

[二、什么是虫洞？](#)

[三、萨根式的问题](#)

[四、虫洞的“创世记”——恼人的因果律](#)

[五、虫洞工程学——负能量的困惑](#)

[六、穿越虫洞——张力的挑战](#)

[七、结语——遥远的天梯](#)

[时间旅行：科学还是幻想？](#)

[一、从《时间机器》讲起](#)

[二、面向未来与重返过去](#)

[三、广义相对论与时间旅行](#)

[四、时间旅行与因果佯谬](#)

[五、凝固长河与平行宇宙](#)

[六、幻想与历史](#)

[第四部分 其他](#)

[从民间“科学家”看科普的局限性](#)

[一、民间“科学家”没有接受过系统的科学训练](#)

[二、民间“科学家”无意接受系统的科学训练](#)

[什么是民间“科学家”](#)

[一、新民科引发的问题](#)

[二、有关民科的几个较具误导性或典型性的观点](#)

[三、民科的定义](#)

[四、民科定义的应用](#)

[学物理能做什么？](#)

[关于普通科普与专业科普](#)

[人名索引](#)

[术语索引](#)

[返回总目录](#)

第一部分 数学

孪生素数猜想^[1]

2003年3月28日，在美国数学研究所（American Institute of Mathematics）位于加州帕洛阿尔托（Palo Alto）的总部，一群来自世界各地的数学家怀着极大的兴趣聆听了圣荷西州立大学（San José State University）数学教授戈德斯通（Daniel Goldston）所做的一个学术报告。在这个报告中，戈德斯通介绍了他和土耳其海峡大学（Boğaziçi University）的数学家伊尔迪里姆（Cem Yıldırım）在证明孪生素数猜想（twin prime conjecture）方面所取得的一个进展。这一进展——如果得到确认的话——将把人们在这一领域中的研究大大推进一步。

那么，什么是孪生素数（twin prime）？什么是孪生素数猜想？戈德斯通和伊尔迪里姆所取得的进展又是什么呢？本文将对这些问题做一个简单介绍。

要介绍孪生素数，首先当然要说一说素数（prime number）这一概念。素数是除了1和自身以外没有其他因子的自然数。在数论中，素数可以说是最纯粹、也最令人着迷的概念。关于素数，一个最简单的事实就是：除了2以外，所有素数都是奇数（因为否则的话，除了1和自身以外还会有一个因子2，从而不满足定义）。由这一简单事实可以得到一个简单推论，那就是：大于2的两个相邻素数之间的最小可能的间隔是2。所谓孪生素数指的就是这种间隔为2的相邻素数，它们之间的距离已经近得不能再近了，就像孪生兄弟一样。不难验证，在孪生素数中，最

小的一对是 $(3, 5)$ ，在100以内则还有 $(5, 7)$ 、 $(11, 13)$ 、 $(17, 19)$ 、 $(29, 31)$ 、 $(41, 43)$ 、 $(59, 61)$ 和 $(71, 73)$ 等另外7对，总计为8对。进一步的验证还表明，随着数字的增大，孪生素数的分布大体上会变得越来越稀疏，寻找孪生素数也会变得越来越困难。

那么，会不会在超过某个界限之后就再也不存在孪生素数了呢？

这个问题让我们联想到素数本身的分布。我们知道，素数本身的分布也是随着数字的增大而越来越稀疏的，因此也有一个会不会在超过某个界限之后就再也不存在的问题。不过幸运的是，早在古希腊时代，著名数学家欧几里得（Euclid）就证明了素数有无穷多个（否则的话——即假如素数没有无穷多个的话——今天的许多数论学家恐怕就得另谋生路了）。长期以来数学家们普遍猜测，孪生素数的情形与素数类似，虽然其分布随着数字的增大而越来越稀疏，总数却是无穷的。这就是与哥德巴赫猜想（Goldbach conjecture）齐名、集令人惊异的表述简单性与令人惊异的证明复杂性于一身的著名猜想——孪生素数猜想。

孪生素数猜想：存在无穷多个素数 p ，使得 $p+2$ 也是素数。

究竟是谁最早明确地提出这一猜想我没有考证过，但1849年法国数学波利尼亚克（Alphonse de Polignac）曾提出过一个猜想：对于任意偶数 $2k$ ，存在无穷多组以 $2k$ 为间隔的素数。这一猜想被称为波利尼亚克猜想（Polignac's conjecture）。对于 $k=1$ ，它就是孪生素数猜想。因此人们有时把波利尼亚克作为孪生素数猜想的提出者。值得一提的是，人们对不同的 k 所对应的素数对的命名是很有趣的： $k=1$ （即间隔为2）的素数对我们已经知道叫做孪生素数； $k=2$ （即间隔为4）的素数对被称为

cousin prime（表兄弟素数），比“孪生”稍远；而 $k=3$ （即间隔为6）的素数对竟被称为sexy prime！这回该相信“书中自有颜如玉”了吧？不过别想歪了，之所以称为sexy prime，其实是因为sex正好是拉丁文中的“6”（因此sexy prime的中文译名乃是毫无联想余地的“六素数”）。

孪生素数猜想还有一个更强的形式，是英国数学家哈代（Godfrey Hardy）和李特伍德（John Littlewood）于1923年提出的，有时被称为哈代-李特伍德猜想（Hardy-Littlewood conjecture）或强孪生素数猜想（strong twin prime conjecture）^[2]。这一猜想不仅提出孪生素数有无穷多组，而且还给出其渐近分布为

$$\pi_2(x) \sim 2C_2 \int_2^x \frac{dt}{(\ln t)^2}$$

其中 $\pi_2(x)$ 表示小于 x 的孪生素数的数目， C_2 被称为孪生素数常数（twin prime constant），其数值为

$$C_2 = \prod_{p \geq 3} \frac{p(p-2)}{(p-1)^2} \approx 0.660\,161\,181\,584\,686\,957\,392\,781\,211\,001\,45\cdots$$

强孪生素数猜想对孪生素数分布的拟合程度可以由表1看出。很明显，拟合程度是相当漂亮的。假如可以拿观测科学的例子来作比拟的话，如此漂亮的拟合几乎能跟英国天文学家亚当斯（John Couch Adams）和法国天文学家勒维耶（Urbain Le Verrier）运用天体摄动规律对海王星位置的预言，以及爱因斯坦（Albert Einstein）的广义相对论对光线引力偏转的预言等最精彩的观测科学成就相媲美，可以算同为理性思维的动人篇章。这种拟合对于纯数学的证明来说虽起不到实质帮助，却大大增强了人们对孪生素数猜想的信心。

表 1

x	孪生素数数目	强孪生素数猜想给出的数目
100 000	1 224	1 249
1 000 000	8 169	8 248
10 000 000	58 980	58 754
100 000 000	440 312	440 368
10 000 000 000	27 412 679	27 411 417

在这里还可以顺便提一下，强孪生素数猜想所给出的孪生素数分布规律可以通过一个简单的定性分析来“得到”^[3]：我们知道，素数定理（prime number theorem）表明对于足够大的 x ，在 x 附近素数的分布密度大约为 $1/\ln(x)$ ，因此两个素数位于宽度为2的区间之内（即构成孪生素数）的概率大约为 $2/\ln^2(x)$ 。这几乎正好就是强孪生素数猜想中的被积函数——当然，两者之间还差了一个孪生素数常数 C_2 ，而这个常数显然正是哈代和李特伍德的功力深厚之处^[4]。

除了强孪生素数猜想与孪生素数实际分布之间的漂亮拟合外，对孪生素数猜想的另一类“实验”支持来自于对越来越大的孪生素数的直接寻找。就像对大素数的寻找一样，这种寻找在很大程度上成为了对计算机运算能力的一种检验。1994年10月30日，这种寻找竟然使人们发现了英特尔（Intel）奔腾（Pentium）处理器浮点除法运算的一个瑕疵（bug），在工程界引起了不小的震动。截至2002年底，人们发现的最大的孪生素数是：

$$(33\,218\,925 \times 2^{169\,690} - 1, 33\,218\,925 \times 2^{169\,690} + 1)$$

这对素数中的每一个都长达51 090位。许多年来这种纪录一直被持续而成功地刷新着，它们对于纯数学的证明来说虽也起不到实质帮助，却同样有助于增强人们对孪生素数猜想的信心[5]。

好了，介绍了这么多关于孪生素数的资料，现在该说说人们在证明孪生素数猜想上所走过的征途了。

迄今为止，在证明孪生素数猜想上的成果大体可以分为两类。第一类是非估算性的，这方面迄今最好的结果是1966年由我国数学家陈景润利用筛法（sieve method）所取得的[6]。陈景润证明了：存在无穷多个素数 p ，使得 $p+2$ 要么是素数，要么是两个素数的乘积。这个结果的形式与他关于哥德巴赫猜想的结果很类似[7]。目前一般认为，由于筛法本身所具有的局限性，这一结果在筛法的范围之内已很难被超越。

证明孪生素数猜想的另一类结果则是估算性的，戈德斯通和伊尔迪里姆所取得的结果就属于这一类。这类结果估算的是相邻素数之间的最小间隔，更确切地说是：

$$\Delta = \liminf_{n \rightarrow \infty} \frac{p_{n+1} - p_n}{\ln p_n}$$

翻译成白话文，这个表达式所定义的是两个相邻素数之间的间隔与其中较小的那个素数的对数值之比在整个素数集合中所取的最小值。很明显，孪生素数猜想要成立， Δ 必须为0。因为孪生素数猜想表明 $p_{n+1} - p_n = 2$ 对无穷多个 n 成立，而 $\ln(p_n) \rightarrow \infty$ ，因此两者之比的最小值对于孪生素数集合——从而对于整个素数集合也——趋于零。不过要注意，

$\Delta=0$ 只是孪生素数猜想成立的必要条件，而不是充分条件。换句话说，如果能证明 $\Delta \neq 0$ ，则孪生素数猜想就被推翻了；但证明了 $\Delta=0$ ，却并不意味着孪生素数猜想一定成立。

对 Δ 最简单的估算来自于素数定理。按照素数定理，对于足够大的 x ，在 x 附近素数出现的几率为 $1/\ln(x)$ ，这表明素数之间的平均间隔为 $\ln(x)$ ，从而 $(p_{n+1} - p_n) / \ln(p_n)$ 给出的其实是相邻素数之间的间隔与平均间隔的比值，其平均值显然为1^[8]。平均值为1，最小值显然是小于等于1，因此素数定理给出 $\Delta \leq 1$ 。

对 Δ 的进一步估算始于哈代和李特伍德。1926年，他们运用圆法（circle method）证明了假如广义黎曼猜想（generalized Riemann hypothesis）成立，则 $\Delta \leq 2/3$ 。这一结果后来被苏格兰数学家兰金（Robert Alexander Rankin）改进为 $\Delta \leq 3/5$ 。但这两个结果都有赖于本身尚未得到证明的广义黎曼猜想，因此只能算是有条件的结果。1940年，匈牙利数学家埃尔德什（Paul Erdős）利用筛法率先给出了一个不带条件的结果： $\Delta < 1$ （即把素数定理给出的结果中的等号部分去掉了）。此后意大利数学家里奇（Giovanni Ricci）于1954年，意大利数学家蓬皮埃利（Enrico Bombieri）、英国数学家达文波特（Harold Davenport）于1966年，以及英国数学家赫克斯利（Martin Huxley）于1977年，分别将 Δ 的估算值推进到了 $\Delta \leq 15/16$ ， $\Delta \leq (2 + \sqrt{3})/8 \approx 0.4665$ ，以及 $\Delta \leq 0.4425$ 。戈德斯通和伊尔迪里姆之前最好的结果则是德国数学家梅尔（Helmut Maier）于1986年得到的 $\Delta \leq 0.2486$ 。

以上这些结果都是在小数点后面做文章，戈德斯通和伊尔迪里姆的结果将这一系列努力大大推进了一步，并且——如果得到确认的话——将在一定意义上终结对 Δ 进行数值估算的长达几十年的漫漫征途。因为

戈德斯通和伊尔迪里姆所证明的结果是 $\Delta=0$ 。当然，如我们前面所述， $\Delta=0$ 只是孪生素数猜想成立的必要条件，而不是充分条件，因此戈德斯通和伊尔迪里姆的结果即便得到确认，离最终证明孪生素数猜想仍有相当的距离，但它无疑将是近十几年来这一领域中最引人注目的结果。

一旦 $\Delta=0$ 被证明，下一个努力方向会是什么呢？一个很自然的方向将是研究 Δ 趋于0的方式。孪生素数猜想要求 $\Delta \sim [\ln(p_n)]^{-1}$ （因为 $p_{n+1}-p_n=2$ 对无穷多个 n 成立）。戈德斯通和伊尔迪里姆的结果所给出的则是 $\Delta \sim [\ln(p_n)]^{-1/9}$ ，两者之间还有不小的差距^[9]。但是看过戈德斯通和伊尔迪里姆手稿的一些数学家认为，戈德斯通和伊尔迪里姆所用的方法还存在改进空间。也就是说，他们的方法还有可能对 Δ 趋于0的方式作出更强的估计。从这个意义上讲，戈德斯通和伊尔迪里姆这一结果的价值不仅仅在于结果本身，更在于它有可能成为一系列未来研究的起点。这种带传承性的系列研究对于数学来说有着双重的重要性，因为一方面，这种研究可能取得的新结果将是对数学的直接贡献；另一方面，这种研究对戈德斯通和伊尔迪里姆的结果会起到反复推敲与核实的作用。现代数学早已超越了一两个评审花一两个小时就可以对一个数学证明做出评判的时代。著名的四色定理（four color theorem）和费马大定理（Fermat's Last Theorem）都曾有过一个证明时隔几年、甚至十几年才被发现错误的例子。因此，一个复杂的数学结果能成为进一步研究的起点，吸引其他数学家的参与，对于最终判定其正确性具有极其正面的意义^[10]。

2003年4月6日写于纽约

2014年9月15日最新修订

[1] 本文撰写于2003年4月，是我的第一篇数学科普，填补了作为本人兴趣主要组成部分之一的数学在我网站的空白。自那以后，本文曾以“补注”形式对若干后续进展作了简单提及，并于2014年9月进行了不改变基本结构的轻微修订。

[2] 确切地说，哈代和李特伍德于1923年所提出的猜想共有两个，分别称为第一哈代-李特伍德猜想（first Hardy-Littlewood conjecture）和第二哈代-李特伍德猜想（second Hardy-Littlewood conjecture）。其中第一哈代-李特伍德猜想又称为 k -tuple猜想（ k -tuple conjecture），它给出了所有形如 $(p, p+2m_1, \dots, p+2m_k)$ （其中 $0 < m_1 < \dots < m_k$ ）的素数 k -tuple的渐进分布。强孪生素数猜想只是 k -tuple猜想的一个特例。

[3] 这种定性分析被澳大利亚数学家陶哲轩（Terence Tao）称为“概率启发式理由”（probabilistic heuristic justification），它不是证明，但对于判断命题成立与否有一定的启示性。

[4] 对孪生素数常数 C_2 也存在“概率启发式理由”，感兴趣的读者可参阅美国数学家查基尔（Don Zagier）的“The First 50 Million Prime Numbers”，Math. Intel. 0, 221-224(1977)。

[5] 截至2011年底，这一纪录已被刷新为了： $(3\,756\,801\,695\,685 \times 2^{666\,669} - 1, 3\,756\,801\,695\,685 \times 2^{666\,669} + 1)$ ，这对素数中的每一个都长达200 700位。

[6] 顺便说一下，美国数学研究所在介绍本文开头所提到的戈德斯通和伊尔迪里姆的结果的简报中提到陈景润时所用的称呼是“伟大的中国数学家陈”（the great Chinese mathematician Chen）。

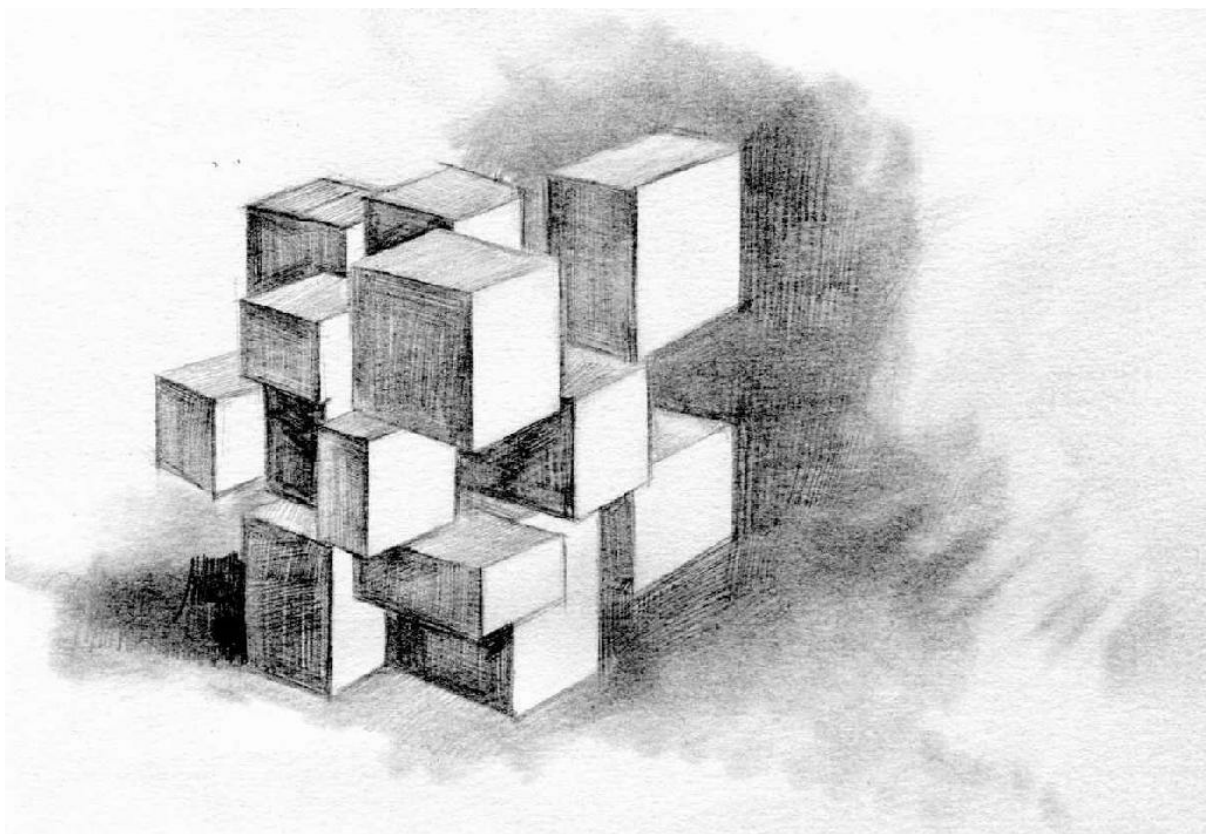
[7] 陈景润关于哥德巴赫猜想的结果——被称为陈氏定理（Chen's theorem）——是：任何足够大的偶数都可以表示成两个数的和，其中一个素数，另一个要么是素数，要么是两个素数的乘积。

[8] 这个“归一”性也正是在 Δ 的表达式中引进 $\ln(p_n)$ 的原因。

[9] 本文发布之后，关于戈德斯通和伊尔迪里姆的工作又有了一些重要的后续发展，其中包括：

2003年4月23日，英国数学家格兰维尔（Andrew Granville）和印度数学家桑德拉拉扬（Kannan Soundararajan）发现了戈德斯通和伊尔迪里姆原始证明中的一个错误，并得到了戈德斯通和伊尔迪里姆的承认；2005年初，戈德斯通和伊尔迪里姆“伙同”匈牙利数学家平兹（János Pintz）“卷土重来”，再次证明了 $\Delta=0$ 。他们所证明的 Δ 的新的渐进行为是： $\Delta \sim [\ln \ln(p_n)]^2 / [\ln(p_n)]^{1/2}$ 。

[10]2013年5月14日，《自然》（*Nature*）等科学杂志及大量中外媒体报道了旅美数学家张益唐在孪生素数猜想研究中所取得的一个重要的新进展，即证明了存在无穷多个素数对，其间隔小于7 000万。这一进展——如果得到确认的话——相当于证明了波利尼亚克猜想至少对某个小于3 500万的 k 成立。用 Δ 来表述的话，则相当于不仅证明了 $\Delta=0$ ，而且给出了与孪生素数猜想所要求的相同的渐进行为： $\Delta \sim [\ln(p_n)]^{-1}$ （不过，这一渐进行为跟 $\Delta=0$ 一样，只是孪生素数猜想成立的必要条件，而不是充分条件）。张益唐的证明用到了戈德斯通、平兹、伊尔迪里姆等人的结果，并于2013年5月21日被《数学年刊》（*Annals of Mathematics*）所接受。张益唐的结果也存在改进空间，截至2014年3月，陶哲轩等数学家已将其中的7 000万这一素数间隔“压缩”到了246。



绘画：张京

魔方与“上帝之数”^[1]

2008年7月，来自世界各地的很多优秀的魔方玩家聚集在捷克共和国（Czech Republic）中部城市帕尔杜比采（Pardubice），参加魔方界的重要赛事：捷克公开赛。在这次比赛中，荷兰玩家阿克斯迪杰克（Erik Akkersdijk）创下了一个惊人的纪录：只用7.08秒就复原了一个颜色被打乱的魔方。无独有偶，在这一年的8月，人们在研究魔方背后的数学问题上也取得了重要进展。在本文中，我们就来介绍一下魔方以及它背后的数学问题。

一、风靡世界的玩具

1974年春天，匈牙利布达佩斯应用艺术学院（Budapest College of Applied Arts）的建筑学教授鲁比克（Ernö Rubik）萌生了一个有趣的念头，那就是设计一个教学工具来帮助学生直观地理解空间几何中的各种转动。经过思考，他决定制作一个由一些小方块组成的，各个面能随意转动的 $3\times 3\times 3$ 的立方体。这样的立方体可以很方便地演示各种空间转动。

这个想法虽好，实践起来却面临一个棘手的问题，即如何才能让这样一个立方体的各个面能随意转动？鲁比克想了很多点子，比如用磁铁或橡皮筋连接各个小方块，但都不成功。那年夏天的一个午后，他在多瑙河畔乘凉，眼光不经意地落在了河畔的鹅卵石上。忽然，他心中闪过一个新的设想：用类似于鹅卵石表面那样的圆形表面来处理立方体的内部结构。这一新设想成功了，鲁比克很快完成了自己的设计，并向匈牙利专利局申请了专利。这一设计就是我们都很熟悉的魔方（magic cube），也叫鲁比克方块（Rubik's cube）^[2]。

6年后，鲁比克的魔方经过一位匈牙利商人兼业余数学家的牵头，打进了西欧及美国市场，并以惊人的速度成为了风靡全球的新潮玩具。在此后的25年间，魔方的销量超过了3亿个。在魔方的玩家中，既有牙牙学语的孩子，也有跨国公司的老总。魔方虽未如鲁比克设想的那样成为一种空间几何的教学工具，却变成了有史以来最畅销的玩具。

魔方之畅销，最大的魔力就在于其数目惊人的颜色组合。一个魔方出厂时每个面各有一种颜色，总共有6种颜色，但这些颜色被打乱后，所能形成的组合数却多达4 325亿亿^[3]（1亿亿= 1×10^{16} ）。如果我们将这

些组合中的每一种都做成一个魔方，这些魔方排在一起，可以从地球一直排到250光年外的遥远星空——也就是说，如果我们在这样一排魔方的一端点上一盏灯，那灯光要在250年后才能照到另一端！如果哪位勤勉的玩家想要尝试所有的组合，哪怕他不吃、不喝、不睡，每秒钟转出10种不同的组合，也要花1 500亿年的时间才能如愿（作为比较，我们的宇宙目前还不到140亿岁）。与这样的组合数相比，广告商们常用的“成千上万”、“数以亿计”、“数以十亿计”等平日里虚张声势、忽悠顾客的形容词反倒变成了难得的谦虚。我们可以很有把握地说，假如不掌握诀窍地随意乱转，一个人哪怕从宇宙大爆炸之初就开始玩魔方，也几乎没有任何希望将一个色彩被打乱的魔方复原。

二、魔方与“上帝之数”

魔方的玩家多了，相互间的比赛自然是少不了的。自1981年起，魔方爱好者们开始举办世界性的魔方大赛，从而开始缔造自己的世界纪录。这一纪录被不断地刷新着，截至2013年，复原魔方的最快纪录已经达到了令人吃惊的5.55秒。当然，单次复原的纪录存在一定的偶然性，为了减少这种偶然性，自2003年起，魔方大赛的冠军改由多次复原的平均成绩来决定^[4]，截至2013年，这一平均成绩的世界纪录为6.54秒。这些纪录的出现，表明魔方虽有天文数字般的颜色组合，但只要掌握窍门，将任何一种给定的颜色组合复原所需的转动次数却很可能并不多。

那么，最少需要多少次转动，才能确保无论什么样的颜色组合都能被复原呢^[5]？这个问题引起了很多人的兴趣。尤其是数学家们的兴趣。这个复原任意组合所需的最少转动次数被数学家们戏称为“上帝之数”（God's number），而魔方这个玩具世界的宠儿则由于这个“上帝之数”而一举侵入了学术界。

要研究“上帝之数”，首先当然要研究魔方的复原方法。在玩魔方的过程中，人们早就知道，将任何一种给定的颜色组合复原都是很容易的，这一点已由玩家们的无数杰出纪录所示范。不过魔方玩家们所用的复原方法是便于人脑掌握的方法，却不是转动次数最少的，因此无助于寻找“上帝之数”。寻找转动次数最少的方法是一个有一定难度的数学问题。当然，这个问题是难不倒数学家的。早在20世纪90年代中期，人们就有了较实用的算法，可以用平均15分钟左右的时间找出复原一种给定的颜色组合的最少转动次数。从理论上讲，如果有人能对每一种颜色组合都找出这样的最少转动次数，那么这些转动次数中最大的一个无疑

就是“上帝之数”了。但可惜的是，“4 325亿亿”这个巨大数字成为了人们窥视“上帝之数”的拦路虎。如果采用上面提到的算法，用上面提到的速度寻找，哪怕用一亿台计算机同时进行，也要用超过1 000万年的时间才能完成。

看来蛮干是行不通的，数学家们于是便求助于他们的老本行：数学。从数学的角度看，魔方的颜色组合虽然千变万化，其实都是由一系列基本操作——即转动——产生的，而且那些操作还具有几个非常简单的特点，比如任何一个操作都有一个相反的操作（比如与顺时针转动相反的操作就是逆时针转动）。对于这样的操作，数学家们的“武器库”中有一种非常有效的工具来对付它，这工具叫做群论（group theory），它比魔方问世早了140多年就已出现了。据说德国数学大师希尔伯特（David Hilbert）曾经表示，学习群论的窍门就是选取一个好的例子。自魔方问世以来，数学家们已经写出了好几本通过魔方讲述群论的书。因此，魔方虽未成为空间几何的教学工具，却在一定程度上可以作为学习群论的“好的例子”。

对魔方研究来说，群论有一个非常重要的优点，就是可以充分利用魔方的对称性。我们前面提到“4 325亿亿”这个巨大数字时，其实有一个疏漏，那就是未曾考虑到魔方作为一个立方体所具有的对称性。由此导致的结果，是那4 325亿亿种颜色组合中有很多其实是完全相同的，只是从不同的角度去看——比如让不同的面朝上或者通过镜子去看——而已。因此，“4 325亿亿”这个令人望而生畏的数字实际上是“注水猪肉”。那么，这“猪肉”中的“水分”占多大比例呢？说出来吓大家一跳：占了将近99%！换句话说，仅凭对称性一项，数学家们就可以把魔方的颜色组合减少两个数量级^[6]。

但减少两个数量级对于寻找“上帝之数”来说还是远远不够的，因为

那不过是将前面提到的1000万年的时间减少为了10万年。对于解决一个数学问题来说，10万年显然还是太长了，而且我们也并不指望真有人能动用一亿台计算机来计算“上帝之数”。数学家们虽然富有智慧，在其他方面却不见得富有，他们真正能动用的也许只有自己书桌上那台计算机。因此为了寻找“上帝之数”，人们还需要更巧妙的思路。幸运的是，群论这一工具的威力远不只是用来分析像立方体的对称性那样显而易见的东西，在它的帮助下，更巧妙的思路很快就出现了。

三、寻找“上帝之数”

1992年，德国数学家科先巴（Herbert Kociemba）提出了一种寻找魔方复原方法的新思路^[7]。他发现，在魔方的基本转动方式中，有一部分可以自成系列，通过这部分转动可以形成将近200亿种颜色组合^[8]。利用这200亿种颜色组合，科先巴将魔方的复原问题分解成了两个步骤：第一步是将任意一种颜色组合转变为那200亿种颜色组合之一，第二步则是将那200亿种颜色组合复原。如果我们把魔方的复原比作是让一条汪洋大海中的小船驶往一个固定目的地，那么科先巴提出的那200亿种颜色组合就好比是一片特殊水域——一片比那个固定目的地大了200亿倍的特殊水域。他提出的两个步骤就好比是让小船首先驶往那片特殊水域，然后从那里驶往那个固定目的地。在汪洋大海中寻找一片巨大的特殊水域，显然要比直接寻找那个小小的目的地容易得多，这就是科先巴新思路的巧妙之处。

但即便如此，要用科先巴的新思路对“上帝之数”进行估算仍不是一件容易的事。尤其是，要想进行快速计算，最好是将复原那200亿种颜色组合的最少转动次数（这相当于是那片特殊水域的“地图”）存储在计算机的内存中，这大约需要300兆（300MB）的内存。300兆在今天看来是一个不太大的数目，但在科先巴提出新思路的年代，普通计算机的内存连它的十分之一都远远不到。因此直到3年之后，才有人利用科先巴的新思路给出了第一个估算结果。此人名叫里德（Michael Reid），是美国中佛罗里达大学（University of Central Florida）的数学家。1995年，里德通过计算发现，最多经过12次转动，就可以将魔方的任意一种颜色组合转变为科先巴新思路中那200亿种颜色组合之一；而最多经过18次转动，就可以将那200亿种颜色组合中的任意一种复原。这表明，

最多经过 $12+18=30$ 次转动，就可以将魔方的任意一种颜色组合复原。

在得到上述结果后，里德很快对自己的估算作了改进，将结果从30减少为了29，这表明“上帝之数”不会超过29。此后随着计算机技术的发展，数学家们对里德的结果又作出了进一步改进，但进展并不迅速。直到11年后的2006年，奥地利开普勒大学（Johannes Kepler University）符号计算研究所（Research Institute for Symbolic Computation）的博士生拉杜（Silviu Radu）才将结果推进到了27。第二年（即2007年），美国东北大学（Northeastern University）的计算机科学家孔克拉（Dan Kunkle）和库伯曼（Gene Cooperman）又将结果推进到了26，他们的工作采用了并行计算系统，所用的最大存储空间高达700万兆（ $7\times 10^6\text{MB}$ ），所耗的计算时间则长达8 000小时（相当于将近一年的24小时不停歇计算）。

这些计算表明，“上帝之数”不会超过26。但是，所有这些计算的最大优点——即利用科先巴新思路中那片特殊水域——同时也是它们最致命的弱点，因为它们给出的复原方法都必须经过那片特殊水域。可事实上，很多颜色组合的最佳复原方法根本就不经过那片特殊水域，比如紧邻目的地，却恰好不在特殊水域中的任何小船，显然都没必要像中国大陆和台湾之间的直航包机一样，故意从那片特殊水域绕一下才前往目的地。因此，用科先巴新思路得到的复原方法未必是最佳的，由此对“上帝之数”所做的估计也极有可能是高估。

可是，如果不引进科先巴新思路中的特殊水域，计算量又实在太太大，怎么办呢？数学家们决定采取折中手段，即扩大那片特殊水域的“面积”。因为特殊水域越大，最佳复原路径恰好经过它的可能性也就越大（当然，计算量也会有相应的增加）。2008年，研究“上帝之数”长

达15年之久的计算机高手罗基奇（Tomas Rokicki）运用了相当于将科先巴新思路中的特殊水域扩大几千倍的巧妙方法，在短短几个月的时间内对“上帝之数”连续发动了四次猛烈攻击，将它的估计值从25一直压缩到了22——这就是本文开头提到的人们在研究魔方背后的数学问题上取得的重要进展。罗基奇的计算得到了电影特效制作商索尼图形图像运作公司（Sony Pictures Imageworks）的支持，这家曾为《蜘蛛人》（*Spider-Man*）等著名影片制作特效的公司向罗基奇提供了相当于50年不停歇计算所需的计算机资源。

由此我们进一步知道，“上帝之数”一定不会超过22。但是，罗基奇虽然将科先巴新思路中的特殊水域扩展得很大，终究仍有一些颜色组合的最佳复原方法是无需经过那片特殊水域的，因此，“上帝之数”很可能比22更小。那么，它究竟是多少呢？种种迹象表明，它极有可能是20。这是因为，人们在过去这么多年的所有努力——其中包括罗基奇直接计算过的大约4 000万亿种颜色组合——中，都从未遇到过任何必须用20次以上转动才能复原的颜色组合，这表明“上帝之数”很可能不大于20；另一方面，人们已经发现了几万种颜色组合，它们必须要用20次转动才能复原，这表明“上帝之数”不可能小于20。将这两方面综合起来，数学家们普遍相信，“上帝之数”的真正数值就是20。

2010年8月，这个游戏与数学交织而成的神秘的“上帝之数”终于水落石出：研究“上帝之数”的“元老”科先巴、“新秀”罗基奇，以及另两位合作者——戴维森（Morley Davidson）和德里奇（John Dethridge）——宣布了对“上帝之数”是20的证明^[9]。他们的证明得到了谷歌公司（Google）提供的相当于英特尔（Intel）四核心处理器35年不停歇计算所需的计算机资源。

因此，现在我们可以用数学特有的确定性来宣布“上帝之数”的数值

了，那就是：20。

2008年11月2日写于纽约

2014年9月18日最新修订

[1]本文曾发表于《科学画报》2008年第12期（上海科学技术出版社出版）。

[2]“魔方”是鲁比克自己为这一设计所取的名字，“鲁比克方块”则是美国玩具公司Ideal Toys所取的名字。在西方国家，鲁比克方块这一名称更为流行，在中国，则是魔方这一名称更为流行。另外要提醒读者的是，魔方有很多种类，本文介绍的3×3×3魔方只是其中最常见的一种。

[3]具体的计算是这样的：在组成魔方的小立方体中，有8个是顶点，它们之间有8!种置换；这些顶点每个有3种颜色，从而在朝向上有 3^7 种组合（由于结构所限，魔方的顶点只有7个能有独立朝向）。类似的，魔方有12个小立方体是边，它们之间有 $12!/2$ 种置换（之所以除以2，是因为魔方的顶点一旦确定，边的置换就只有一半是可能的）；这些边每个有两种颜色，在朝向上有 2^{11} 种组合（由于结构所限，魔方的边只有11个能有独立朝向）。因此，魔方的颜色组合总数为 $8! \times 3^7 \times 12! \times 2^{11}/2 = 43\ 252\ 003\ 274\ 489\ 856\ 000$ ，即大约4 325亿亿。另外值得一提的是，倘若我们允许将魔方拆掉重组，则前面提到的结构限定将不复存在，它的颜色组合数将多达51 900亿亿种。不过颜色组合数的增加并不意味着复原的难度变大，魔方结构对颜色组合数的限制实际上正是使魔方的复原变得困难的主要原因。举个例子来说，26个英文字母在相邻字母的交换之下共有约400亿亿亿种组合，远远多于魔方颜色的组合数，但通过相邻字母的交换将随意排列的26个英文字母复原成从A到Z的初始排列却非常简单。

[4]确切地说是取5次尝试中居中的3次成绩的平均值。

[5]为了使这一问题有意义，当然首先要定义什么是转动。在对魔方的数学研究中，转动是指将魔方的任意一个（包含9个小方块的）面沿顺时针或逆时针方向转动90°或180°，对每个面来

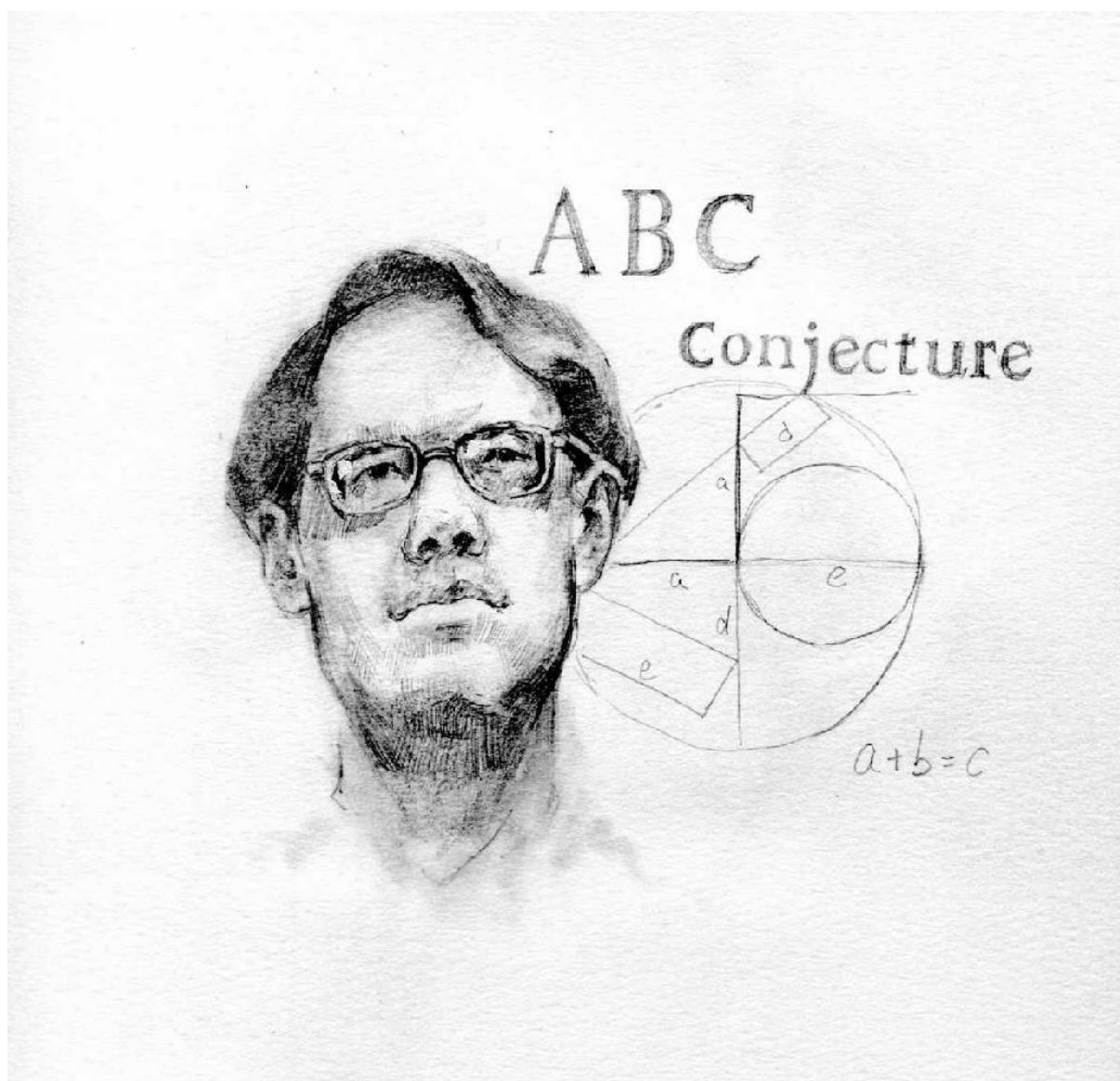
说，这样的转动共有3种。（请读者想一想，为什么不是4种？）由于魔方有6个面，因此它的基本转动方式共有18种。

[6]确切地说，是减少至 $1/96$ ，或45亿亿种组合。

[7]科先巴的新思路是本文介绍的一系列计算研究的起点，但并不是最早的魔方算法。早在1981年，目前在美国田纳西大学（University of Tennessee），当时在伦敦南岸大学（London South Bank University）的数学家西斯尔斯韦特（Morwen Thistlethwaite）就提出了一种算法，被称为西斯尔斯韦特算法（Thistlethwaite algorithm）。西斯尔斯韦特算法可保证通过不超过52次转动复原魔方的任意一种颜色组合（相当于证明了上帝之数不超过52），在科先巴新思路问世之前的1991年，这一数字曾被压缩到42。

[8]确切地说，是18种基本转动方式中有10种自成系列，由此形成的颜色组合共有 $8! \times 8! \times 4! / 2$ （约195亿）种。

[9]他们所宣布的证明完成时间为2010年7月。



绘画：张京

ABC猜想浅说^[1]

由前三个英文字母拼合而成的“ABC”一词据说自13世纪起便见诸文献了，含义为“入门”。这些年随着英文在中国的流行，该词在中文世界里也夺得了一席之地，出现在了很多人图书的书名中，大有跟中文词“入门”一较高下之势。不过，倘若你在数学文献中看到一个以“ABC”命名的猜想——“ABC猜想”（ABC conjecture），千万不要以为那是一个“入门”级别的猜想。事实上，这一猜想在公众知名度方面或许尚处于“入门”阶段，以难度和地位而论却绝不是“入门”级别的。

在本文中，我们将对这一并非“入门”级别的猜想做一个“入门”级别的介绍。

一、什么是ABC猜想？

在介绍之前，让我们先回忆一下中小学数学中的两个简单概念。其中第一个概念是素数（prime number）。我们知道，很多正整数可以分解为其他——即不同于它自己的——正整数的乘积，比如 $9=3\times 3$ ， $231=3\times 7\times 11$ ，等等。但也有一些正整数不能这么分解，比如13，29等。这后一类正整数——1除外——就是所谓的素数。素数是一个被称为“数论”（number theory）的数学分支中的核心概念，其地位常被比喻为物理学中的原子（atom），因为与物理学中物质可以分解为原子相类似，数学中所有大于1的正整数都可以分解为素数的乘积（素数本身被视为是自己的分解）^[2]。第二个概念则是互素（co-prime）。两个正整数如果其素数分解中不存在共同的素数，就称为是互素的，比如 $21=3\times 7$ 和 $55=5\times 11$ 就是互素的^[3]。

有了这两个简单概念，我们就可以介绍ABC猜想了。ABC猜想针对的是满足两个简单条件的正整数组（A,B,C）^[4]。其中第一个条件是A和B互素，第二个条件是 $A + B = C$ 。显然，满足这种条件的正整数组——比如（3，8，11）、（16，17，33）……——有无穷多个（请读者自行证明）。为了引出ABC猜想，让我们以（3，8，11）为例，做一个“三步走”的简单计算：

（1）将A、B、C乘起来（结果是 $3\times 8\times 11=264$ ）；

（2）对乘积进行素数分解（结果是 $264=2^3\times 3\times 11$ ）；

（3）将素数分解中所有不同的素数乘起来（结果是 $2\times 3\times 11=66$ ）。

现在，让我们将 A 、 B 、 C 三个数字中较大的那个（即 C ）与步骤3的结果比较一下。我们发现后者大于前者（因为后者为66，前者为11）。读者可以对上面所举的另一个例子——即（16，17，33）——也试一下，你会发现同样的结果。如果随便找一些其他例子，你也很可能发现同样的结果。

但你若因此以为这是规律，那就完全错了，因为它不仅不是规律，而且有无穷多的反例。比如（3，125，128）就是一个反例（请读者自行验证）。但是，数学家们猜测，如果把步骤3的结果放大成它的一个大于1的幂，那个幂哪怕只比1大上一丁点儿（比如1.000 000 000 01），情况就有可能大不一样。这时它虽仍未必保证能够大于三个数字中较大的那个（即 C ），但反例的数目将由无穷变为有限。这个猜测就是所谓的ABC猜想^[5]，它是由英国数学家麦瑟尔（David Masser）和法国数学家厄斯特勒（Joseph Oesterlé）于20世纪80年代中期彼此独立地提出的。“ABC”这个毫无创意的名字——大家可能猜到了——则是来自把猜想中涉及到的三个数字称为 A 、 B 、 C 的做法，而非“入门”之意。

与数学猜想大家庭中的著名成员，如黎曼猜想（Riemann hypothesis）、哥德巴赫猜想（Goldbach conjecture）、孪生素数猜想（twin prime conjecture），以及（已被证明了的）曾经的费马猜想（Fermat conjecture）、四色猜想（four-color conjecture）等相比，ABC猜想的“资历”是很浅的（其他那些猜想都是百岁以上的“老前辈”），公众知名度也颇有不及，但以重要性而论，则除黎曼猜想外，上述其他几个猜想都得退居其后。

二、ABC猜想为什么重要？

ABC猜想有一个在普通人看来并不奥妙的特点，就是将整数的加法性质（比如 $A+B=C$ ）和乘法性质（比如素数概念——因为它是由乘法性质所定义的）交互在了一起。不过，数学家们早就知道，由这两种本身很简单的性质交互所能产生的复杂性是近乎无穷的。数论中许多表述极为浅显，却极难证明的猜想（或曾经的猜想），比如前面提到的哥德巴赫猜想、孪生素数猜想、费马猜想等都具有这种加法性质和乘法性质相交互的特性。数论中一个很重要的分支——旨在研究整系数代数方程的整数解的所谓丢番图分析（Diophantine analysis）——更是整个分支都具有这一特性。丢番图分析的困难性是颇为出名的，著名德国数学家希尔伯特（David Hilbert）曾乐观地希望能找到其“一揽子”的解决方案，可惜这个被称为希尔伯特第十问题的希望后来落了空，被证明是不可能实现的（对这一点感兴趣的读者可参阅拙作《小楼与大师：科学殿堂的人和事》中的《希尔伯特第十问题漫谈》一文）。与希尔伯特的乐观相反，美国哥伦比亚大学（Columbia University）的数学家戈德菲尔德（Dorian Goldfeld）曾将丢番图分析比喻为飞蝇钓（fly-fishing）——那是发源于英国贵族的一种特殊的钓鱼手法，用甩出去的诱饵模拟飞蝇等昆虫的飞行姿态，以吸引凶猛的掠食性鱼类。飞蝇钓的特点是技巧高、难度大、成功率低，而且只能一条一条慢慢地钓——象征着丢番图分析只能一个问题一个问题慢慢地啃，而无法像希尔伯特所希望的那样“一揽子”地解决掉。

但是，与交互了加法性质和乘法性质的其他猜想或问题不同的是，ABC猜想这个从表述上看颇有些拖泥带水（因为允许反例）的猜想似乎处于某种中枢地位上，它的解决将直接导致一大类其他猜想或问题的解

决。拿丢番图分析来说，戈德菲尔德就表示，假如ABC猜想能被证明，丢番图分析将由飞蝇钓变为最强力——乃至野蛮——的炸药捕鱼，一炸就是一大片，因为ABC猜想能“将无穷多个丢番图方程转变为单一数学命题”。这其中最引人注目的“战利品”将是曾作为猜想存在了300多年，一度被《吉尼斯世界纪录》（*Guinness Book of World Records*）称为“最困难数学问题”的费马猜想。这个直到1995年才被英国数学家怀尔斯（Andrew Wiles）以超过100页的长篇论文所解决的猜想在ABC猜想成立的前提下，将只需不到一页的数学推理就能确立^[6]。其他很多长期悬而未决的数学猜想或问题也将被“一锅端”。这种与其他数学命题之间的紧密联系是衡量一个数学命题重要性的首要“考评”指标，ABC猜想在这方面无疑能得高分——或者用戈德菲尔德的话说，是“丢番图分析中最重要的未解决问题”，“是一种美丽”。

ABC猜想的重要性吸引了很多数学家的兴趣，但它的艰深迟滞了取得进展的步伐。截至2001年，数学家们在这一猜想上取得的最好结果乃是将上述步骤3的结果放大成它的某种指数函数^[7]。由于指数函数的大范围增长速度远比幂函数快得多，由它来保证其大于A、B、C三个数字中较大的那个（即C）当然要容易得多（相应地，命题本身则要弱得多）。

除上述理论结果外，自2006年起，由荷兰莱顿大学（Leiden University）的数学系牵头，一些数学和计算机爱好者建立了一个名为ABC@Home的分布式计算（distributed computing）系统，用以寻找ABC猜想所允许的反例。截至2014年4月，该系统已经找到了超过2 380万个反例，而且还在继续增加着。不过，与这一系统的著名“同行”——比如寻找外星智慧生物的SETI以及计算黎曼 ζ 函数非平凡零点的已经关闭了的ZetaGrid——不同的是，ABC@Home是既不可能证明，也不可能

否定ABC猜想的（因为ABC猜想本就允许数量有限的反例）。从这个意义上讲，ABC@Home的建立更多地只是出于对具体反例——尤其是某些极端情形下的反例，比如数值最大的反例——的好奇。当然，具体反例积累多了，是否会衍生出有关反例分布的猜想，也是不无趣味的悬念。另外，ABC猜想还有一些拓展版本，比如对某些情形下的反例数目给出具体数值的版本，ABC@Home对那种版本原则上是有否定能力的。

三、ABC猜想被证明了吗？

如前所述，ABC猜想的公众知名度与一些著名猜想相比是颇有不及的。不过，2012年9月初，包括《自然》（*Nature*）、《科学》

（*Science*）在内的一些重量级学术刊物，以及包括《纽约时报》（*New York Times*）在内的许多著名媒体却纷纷撰写或转载了有关ABC猜想的消息，使这一猜想在短时间内着实风光了一番。促成这一风光的是日本数学家望月新一（Shinichi Mochizuki）。2012年8月底，望月新一发表了由四篇长文组成的系列论文的第四篇，宣称证明了包括ABC猜想在内的若干重要猜想。这一宣称被一些媒体称为是能与1993年怀尔斯宣称证明了费马猜想，以及2002年佩雷尔曼（Grigory Perelman）宣称证明了庞加莱猜想（Poincaré conjecture）相提并论的事件。

由于这一原因，我应约撰写本文时，约稿编辑曾希望我能找认识望月新一的华人数学家聊聊，挖出点独家新闻来。可惜我不得不有负此托了，因为别说是我，就连《纽约时报》等擅挖材料的重量级媒体在报道望月新一其人时，也基本没能超出他在自己网站上公布的信息。

按照那些信息，望月新一1969年3月29日出生于日本东京，16岁（即1985年）进入美国普林斯顿大学（Princeton University）就读本科，三年后进入研究生院，师从著名德国数学家、1986年菲尔茨奖（Fields Medal）得主法尔廷斯（Gerd Faltings），23岁（即1992年）获得数学博士学位。此后，他先是“海归”成京都大学（Kyoto University）数理解析研究所（Research Institute for Mathematical Sciences）的研究助理（Research Associate），几个月后又前往美国哈佛大学从事了近两年的研究，然后重返京都大学。2002年，33岁的望月新一成为了京都大学

数理解析研究所的教授。望月新一的学术声誉颇佳，曾获得过日本学术奖章（Japan Academy Medal）等荣誉。

有关望月新一其人的信息大体就是这些，但读者不必过于失望，因为望月新一所宣称的对ABC猜想的证明虽引起了很大关注，离公认还颇有距离，因此目前恐怕还未到挖掘其生平的最佳时机。事实上，在ABC猜想并不漫长的历史中，这并不是第一次有人宣称解决了这一猜想。

2007年，法国数学家施皮罗（Lucien Szpiro）就曾宣称解决了ABC猜想。施皮罗的学术声誉不在望月新一之下，不仅是领域内的专家，其工作甚至间接促成了ABC猜想的提出。但是，人们很快就在他的证明中发现了漏洞。这种宣称解决了一个重大数学猜想，随后却被发现漏洞的例子在数学史上比比皆是。因此，任何证明从宣称到公认，必须经过同行的严格检验。这一检验视证明的复杂程度而定，可长可短。不过对于望月新一的“粉丝”来说，恐怕得有长期等待的心理准备，因为望月新一那四篇论文的总长度超过了500页，几乎是怀尔斯证明费马猜想的论文长度的四倍！更糟糕的是，望月新一的证明采用了他自己发展起来的数学工具，这种工具据说是对以抽象和艰深著称的1966年菲尔兹奖得主格罗滕迪克（Alexander Grothendieck）的某些代数几何方法的推广，除他本人外，数学界并无第二人通晓^[8]。就连研究方向与望月新一相近的英国牛津大学（University of Oxford）的韩国数学家金明迥（Minhyong Kim）都表示，“我甚至无法对[望月新一的]证明给出一个专家概述，因为我并不理解它”，“仅仅对局势有一个一般了解也得花费一段时间”。

美国威斯康星大学（University of Wisconsin）的数学家艾伦伯格（Jordan Ellenberg）则表示阅读望月新一的论文“仿佛是在阅读外星人的东西”（reading something from outer space）。2006年菲尔茨奖得主、澳大利亚数学家陶哲轩（Terence Tao）也表示“现在对这一证明有可能正确还是错误做出评断还为时过早”。

像望月新一那样宣称用自创的数学工具证明著名数学猜想的事例在数学界也是有先例的。2004年，美国普渡大学（Purdue）的数学教授德布朗基（Louis de Branges）宣称证明了著名的黎曼猜想，他所用的也是自创的数学工具。不过德布朗基在数学界的声誉和口碑均极差，加之年事已高（七旬老汉），其宣称遭到了数学界的冷淡对待^[9]。与之不同的是，望月新一却不仅有良好的学术声誉，精力和研究能力也尚处于巅峰期。用陶哲轩的话说，望月新一“与佩雷尔曼和怀尔斯类似”，“是一个多年来致力于解决重要问题，在领域内享有很高声誉的第一流数学家”。有鉴于此，数学界不仅对望月新一的证明给予了重视，对他自创的方法也表示了兴趣，比如美国斯坦福大学（Stanford University）的数学家康拉德（Brian Conrad）就表示“激动人心之处不仅在于[ABC]猜想有可能已被解决，而且在于他[望月新一]必须引入的技巧和洞见应该是解决未来数论问题的非常有力的工具”。戈德菲尔德也认为“望月新一的证明如果成立，将是21世纪数学最惊人的成就”。

在这种兴趣的驱动下，一些数学家已经开始对望月新一的证明展开检验与讨论，比如著名数学讨论网站Math Overflow就已出现了一些有金明迥、陶哲轩等一流数学家参与的认真讨论。不过，检验过程何时才能完成，目前还不得而知，检验的结果如何，更是无从预料。证明得到公认固然是很多人乐意见到的，但一个长达500多页的证明存在漏洞也是完全可能的，当年怀尔斯对费马猜想的“只有”100多页的证明，其早期版本就存在过漏洞，经过一年多的时间才得以弥补。不过，无论望月新一的证明是否成立，不少数学家对ABC猜想本身的成立倒是都抱有乐观态度，这一方面是因为能因这一猜想的成立而得到证明的很多数学命题（比如如今被称为费马大定理的费马猜想）已经通过其他途径得到了证明，从而表明ABC猜想的成立与数学的其他部分有很好的相容性（著名的黎曼猜想也有这样的特点）。另一方面，ABC猜想还得到了一些启发

性观点的支持，比如陶哲轩就从所谓的“概率启发式理由”（probabilistic heuristic justification）出发，预期ABC猜想应该成立^[10]。

当然，信心和预期取代不了证明。望月新一证明的命运将会如何？ABC猜想究竟被证明了没有？都将有待时间来回答^[11]。

2012年10月14日写于纽约

2014年10月1日最新修订

[1]本文是应《南方周末》约稿而写的“ABC猜想”简介，曾以《望月“摘月”》为标题发表于2012年10月25日（发表稿经编辑改动，系删节版）。本文的完整版发表于《数学文化》2014年11月刊。

[2]不仅如此，这样的分解还可以被证明是唯一的，这被称为算术基本定理（fundamental theorem of arithmetic）。

[3]对这一定义还有一个小小的补充，即1被定义为与所有正整数都互素。

[4]为了简单起见，我们的介绍是针对正整数的，但ABC猜想其实也可以针对整数进行表述，两者并无实质差别。我们将后者留给感兴趣的读者去做。

[5]这里可以略作一点补充：步骤3的结果因不含任何素数因子的平方，被称为A、B、C三个数字乘积的“无平方部分”（square-free part），简记为 $\text{sfp}(ABC)$ ——不过要注意的是，这一记号在某些文献中有不同含义，与本文含义相一致的另一种记号为 $\text{rad}(ABC)$ 。用这一记号，ABC猜想可以表述为“对任意给定的 $n > 1$ ，只有有限多组 (A, B, C) 满足 $\text{sfp}(ABC)^n < C$ ”（当然，别忘了A和B互素及 $A+B=C$ 这两个条件）。这一表述通常见诸科普介绍，在专业文献中ABC猜想往往被表述为“对任意给定的 $n > 1$ ， $\text{sfp}(ABC)^n/C$ 的下界大于零”。感兴趣的读者不妨由“科普表

述”出发，证明一下“专业表述”（不过要提醒读者的是：相反方向的证明，即由“专业表述”证明“科普表述”，并不是轻而易举的）。另外要说明的是，正文提到的所谓ABC猜想所允许的“反例”乃是“科普表述”特有的提法，意指满足 $\text{sqp}(ABC)^n < C$ 的那有限多组 (A, B, C) ，在“专业表述”中是没有所谓“反例”的提法的。

[6]这个关于在ABC猜想成立的前提下，费马猜想将只需“不到一页的数学推理就能确立”（establishing in less than a page of mathematical reasoning）的不无夸张的说法出自美国数学协会（Mathematical Association of America）的出版主管、著名美国数学科普作家彼得森（Ivars Peterson）。不过，该说法虽然夸张，却并非完全“忽悠”。为了说明这一点，并作为对如何由ABC猜想证明其他命题的演示，我们在这里介绍一个“不到一页的数学推理”：假设费马猜想不成立，即存在互素的（这点请读者自行证明）正整数 x, y, z 使得 $x^k + y^k = z^k$ ($k > 2$)。则由前一条注释给出的ABC猜想的“专业表述”可知（取 $n = 7/6$ ）： $\text{sqp}(x^k y^k z^k)^{7/6} / z^k > \varepsilon$ ($\varepsilon > 0$)。由于 $\text{sqp}(x^k y^k z^k) = \text{sqp}(xyz) \leq xyz < z^3$ ，因此 $z^{3, 5-k} > \varepsilon$ 。显然，对所有 $k \geq 4$ ，只有小于（由 ε 决定的）某个数值的有限多个 z 能满足该不等式，而且当 k 大于（由 ε 决定的）某个数值后，将不会有任何 z 满足该不等式。这表明，对所有 $k \geq 4$ ，费马猜想的反例即便有也只能有有限多个，而且 k 大到一定程度后将不再有反例。因此，证明费马猜想就变成了证明 $k=3$ 的情形（这在两百多年前就已完成），以及通过数值验证排除总数有限的反例。这虽然并非“不到一页的数学推理”就能确立的，比起怀尔斯的证明来毕竟是直截了当多了。倘若历史走的是不同的路径，费马是在ABC猜想被证明之后才提出的费马猜想，他那句戏剧性的“我发现了一个真正出色的证明，可惜页边太窄写不下来”倒是不无成立之可能。

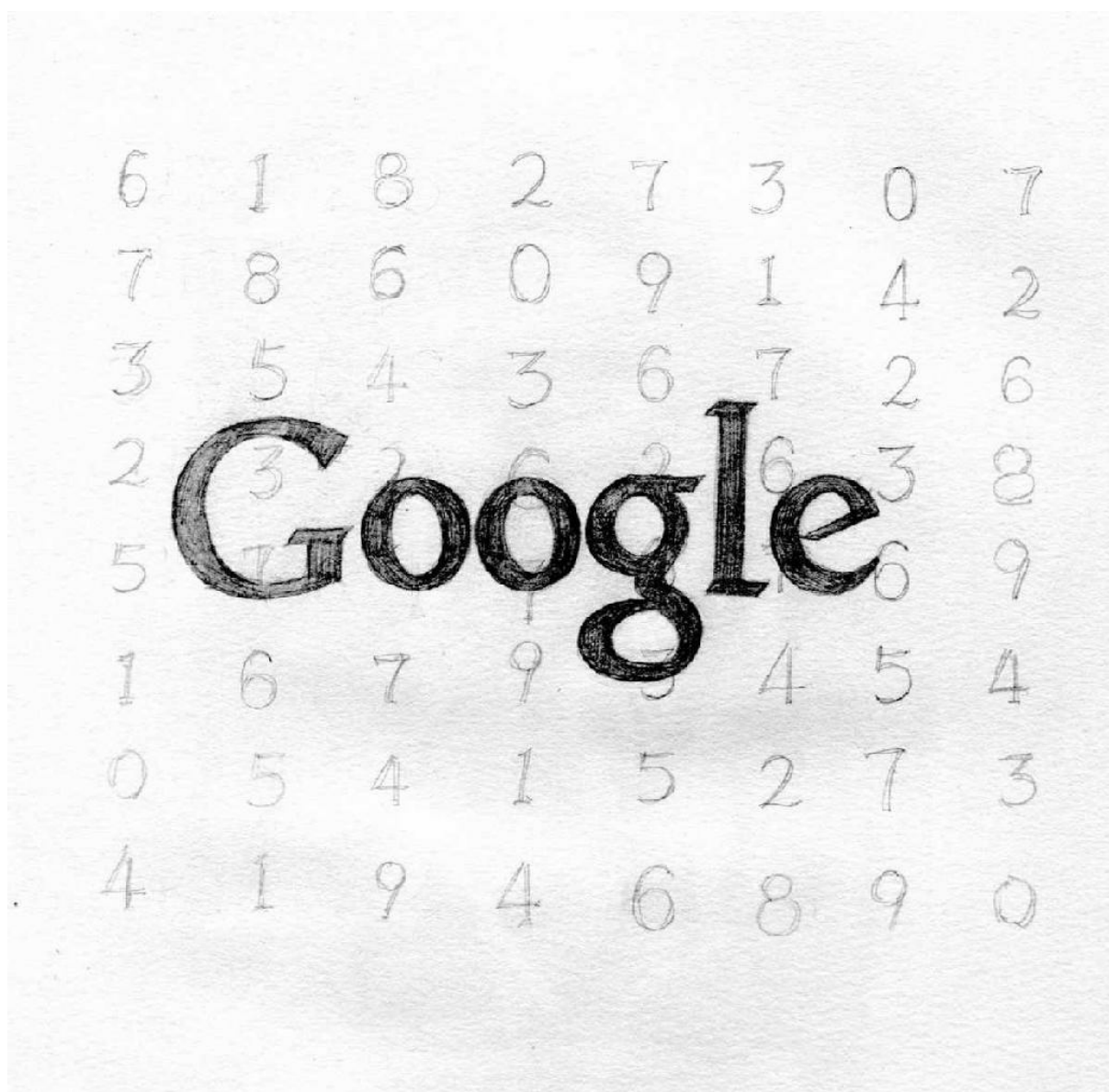
[7]具体地说，截至2001年，这方面的最好结果是 $\exp[K \cdot \text{sqp}(ABC)^{1/3+\varepsilon}] / C > 1$ ，其中 K 是与 ε 有关（但与 A, B, C 无关）的常数。

[8]望月新一自创的那种数学工具被称为inter-universal Teichmüller theory或inter-universal geometry。他在其网站上则称自己为Inter-universal Geometer。

[9]对此事感兴趣的读者可参阅拙作《黎曼猜想漫谈》的第35章。

[10]陶哲轩的“概率启发式理由”的要点是将数论命题——比如一个数是素数——视为概率性命题，并利用概率工具来猜测数学命题的成立与否。这种做法的一个例子是对强孪生素数猜想成立的猜测（参阅收录于本书的拙作“孪生素数猜想”所介绍的有关该猜想的“简单的定性分析”）。

[11]望月新一的证明发布至今已两年多，这期间美国耶鲁大学（Yale University）的数学系研究生季米特洛夫（Vesselin Dimitrov）及斯坦福大学（Stanford University）的数学家文卡塔斯（Akshay Venkatesh）曾写信向他指出过一个错误。望月新一承认了错误，但表示那是一个不影响结论的小错误。此后，他数度更新了自己的论文，截至本文修订之日（2014年10月1日），他更新后的四篇论文总长度超过了550页，最近一次更新的日期则为2014年9月15日。



绘画：张京

谷歌背后的数学^[1]

一、引言

在如今这个互联网时代，有一家公司家喻户晓——它自1998年问世以来，在极短的时间内就声誉鹊起，不仅超越了所有竞争对手，而且彻底改观了整个互联网的生态。这家公司就是当今互联网上的第一搜索引擎：谷歌（Google）。

在这样一家显赫的公司背后，自然有许许多多商战故事，也有许许多多成功因素。但与普通商战故事不同的是，在谷歌的成功背后起着最关键作用的却是一个数学因素。

本文要谈的就是这个数学因素。

谷歌作为一个搜索引擎，它的核心功能顾名思义，就是网页搜索。说到搜索，我们都不陌生，因为那是凡地球人都会的技能。我们在字典里查个生字，在图书馆里找本图书，甚至在商店里寻一种商品，等等，都是搜索。只要稍稍推究一下，我们就会发现那些搜索之所以可能，并且人人都会，在很大程度上得益于以下三条：

（1）搜索对象的数量较小——比如一本字典收录的字通常只有一两万个，一家图书馆收录的不重复图书通常不超过几十万种，一家商店的商品通常不超过几万种，等等。

（2）搜索对象具有良好的分类或排序——比如字典里的字按拼音排序，图书馆里的图书按主题分类，商店里的商品按品种或用途分类，等等。

（3）搜索结果的重复度较低——比如字典里的同音字通常不超过几十个，图书馆里的同名图书和商店里的同种商品通常也不超过几十种，等等。

但互联网的鲜明特点却是以上三条无一满足。事实上，即便在谷歌问世之前，互联网上的网页总数就已超过了诸如图书馆藏书数量之类传统搜索对象的数目。而且这还只是冰山一角，因为与搜索图书时单纯的书名搜索不同，互联网上的搜索往往是对网页内容的直接搜索，这相当于将图书里的每一个字都变成了搜索对象，由此导致的数量才是真正惊人的，它不仅直接破坏了上述第一条，而且连带破坏了二、三两条。在互联网发展的早期，像雅虎（Yahoo）那样的门户网站曾试图为网页建立分类系统，但随着网页数量的激增，这种做法很快就“挂一漏万”了。而搜索结果的重复度更是以快得不能再快的速度走向失控。这其实是可以预料的，因为几乎所有网页都离不开几千个常用词，因此除非搜索生僻词，否则出现几十万、几百万、甚至几千万条搜索结果都是不足为奇的。

互联网的这些“不良特点”给搜索引擎的设计带来了极大的挑战。而在这些挑战之中，相对来说，对一、二两条的破坏是比较容易解决的，因为那主要是对搜索引擎的存储空间和计算能力提出了较高要求，只要有足够多的钱来买“装备”，这些都还能算是容易解决的——套用电视连续剧《蜗居》中某贪官的台词来说，“能用钱解决的问题就不是大问题”。但对第三条的破坏却要了命了，因为无论搜索引擎的硬件如何强大，速度如何快捷，要是搜索结果有几百万条，那么任何用户想从其

中“海选”出自己真正想要的东西都是几乎不可能的。这一点对早期搜索引擎来说可谓是致命伤，而且它不是用钱就能解决的问题。

这致命伤该如何治疗呢？药方其实很简单，那就是对搜索结果进行排序，把用户最有可能需要的网页排在最前面，以确保用户能很方便地找到它们。但问题是：网页的水平千差万别，用户的喜好更是万别千差，互联网上有一句流行语叫做：“在互联网上，没人知道你是一条狗（On the Internet, nobody knows you're a dog）。”连用户是人是狗都“没人知道”，搜索引擎又怎能知道哪些搜索结果是用户最有可能需要的，并对它们进行排序呢？

在谷歌主导互联网搜索之前，多数搜索引擎采用的排序方法，是以被搜索词语在网页中的出现次数来决定排序——出现次数越多的网页排在越前面。这个判据不能说毫无道理，因为用户搜索一个词语，通常表明对该词语感兴趣。既然如此，那该词语在网页中的出现次数越多，就越有可能表示该网页是用户所需要的。可惜的是，这个貌似合理的方法实际上却行不大通。因为按照这种方法，任何一个像祥林嫂一样翻来覆去倒腾某些关键词的网页，无论水平多烂，一旦被搜索到，都会立刻“金榜题名”，这简直就是广告及垃圾网页制造者的天堂。事实上，当时几乎没有一个搜索引擎不被“祥林嫂”们所困扰，其中最具讽刺意味的是：在谷歌诞生之前的1997年11月，堪称早期互联网巨子的当时四大搜索引擎在搜索自己公司的名字时，居然只有一个能使之出现在搜索结果的前十名内，其余全被“祥林嫂”们挤跑了。

二、基本思路

正是在这种情况下，1996年初，谷歌公司的创始人，当时还是美国斯坦福大学（Stanford University）研究生的佩奇（Larry Page）和布林（Sergey Brin）开始了对网页排序问题的研究。这两位小伙子之所以研究网页排序问题，一来是导师的建议（佩奇后来称该建议为“我有生以来得到过的最好建议”），二来则是因为他们对这一问题背后的数学产生了兴趣。

网页排序问题的背后有什么样的数学呢？这得从佩奇和布林看待这一问题的思路说起。

在佩奇和布林看来，网页的排序是不能靠每个网页自己来标榜的，无论把关键词重复多少次，垃圾网页依然是垃圾网页。那么，究竟什么才是网页排序的可靠依据呢？出身于书香门第的佩奇和布林（两人的父亲都是大学教授）想到了学术界评判学术论文重要性的通用方法，那就是看论文的引用次数。在互联网上，与论文的引用相类似的显然是网页的链接。因此，佩奇和布林萌生了一个网页排序的思路，那就是通过研究网页间的相互链接来确定排序。具体地说，一个网页被其他网页链接得越多，它的排序就应该越靠前。不仅如此，佩奇和布林还进一步提出，一个网页越是被排序靠前的网页所链接，它的排序就也应该越靠前。这一条的意义也是不言而喻的，就好比一篇论文被诺贝尔奖得主所引用，显然要比被普通研究者所引用更说明其价值。依照这个思路，网页排序问题就跟整个互联网的链接结构产生了关系，正是这一关系使它成为了一个不折不扣的数学问题。

思路虽然有了，具体计算却并非易事，因为按照这种思路，想要知

道一个网页 w_i 的排序，不仅要知道有多少网页链接了它，而且还得知道那些网页各自的排序——因为来自排序靠前网页的链接更有分量。但作为互联网大家庭的一员， w_i 本身对其他网页的排序也是有贡献的，而且基于来自排序靠前网页的链接更有分量的原则，这种贡献与 w_i 本身的排序也有关。这样一来，我们就陷入了一个“先有鸡还是先有蛋”的循环：要想知道 w_i 的排序，就得知道与它链接的其他网页的排序，而要想知道那些网页的排序，却又首先得知道 w_i 的排序。

为了打破这个循环，佩奇和布林采用了一个很巧妙的思路，即分析一个虚拟用户在互联网上的漫游过程。他们假定：虚拟用户一旦访问了一个网页后，下一步将有相同的几率访问被该网页所链接的任何一个其他网页。换句话说，如果网页 w_i 有 N_i 个对外链接，则虚拟用户在访问了 w_i 之后，下一步点击那些链接当中的任何一个的几率均为 $1/N_i$ 。初看起来，这一假设并不合理，因为任何用户都有偏好，怎么可能以相同的几率访问一个网页的所有链接呢？但如果我们考虑到佩奇和布林的虚拟用户实际上是对互联网上全体用户的一种平均意义上的代表，这条假设就不像初看起来那么不合理了。那么网页的排序由什么来决定呢？是由该用户在漫游了很长时间——理论上为无穷长时间——后访问各网页的几率分布来决定，访问几率越大的网页排序就越靠前。

为了将这一分析数学化，我们用 $p_i(n)$ 表示虚拟用户在进行第 n 次浏览时访问网页 w_i 的几率。显然，上述假设可以表述为（请读者自行证明）：

$$p_i(n+1) = \sum_j \frac{p_j(n)p_{j \rightarrow i}}{N_j}$$

这里 $p_{j \rightarrow i}$ 是一个描述互联网链接结构的指标函数（indicator function），其定义是：如果网页 W_j 有链接指向网页 W_i ，则 $p_{j \rightarrow i}$ 取值为1，反之则为0。显然，这条假设所体现的正是前面提到的佩奇和布林的排序原则，因为右端求和式的存在表明与 W_i 有链接的所有网页 W_j 都对 W_i 的排名有贡献，而求和式中的每一项都正比于 p_j ，则表明来自那些网页的贡献与它们的自身排序有关，自身排序越靠前（即 p_j 越大），贡献就越大。

为符号简洁起见，我们将虚拟用户第 n 次浏览时访问各网页的几率合并为一个列向量 \mathbf{p}_n ，它的第 i 个分量为 $p_i(n)$ ，并引进一个只与互联网结构有关的矩阵 \mathbf{H} ，它的第 i 行 j 列的矩阵元为 $H_{ij}=p_{j \rightarrow i}/N_j$ ，则上述公式可以改写为

$$\mathbf{p}_{n+1} = \mathbf{H}\mathbf{p}_n$$

这就是计算网页排序的公式。

熟悉随机过程理论的读者想必看出来了，上述公式描述的是一种马尔可夫过程（Markov process），而且是其中最简单的一类，即所谓的平稳马尔可夫过程（stationary Markov process）^[2]，而 \mathbf{H} 则是描述马尔可夫过程中的转移概率分布的所谓转移矩阵（transition matrix）。不过普通马尔可夫过程中的转移矩阵通常是随机矩阵（stochastic matrix），即每一列的矩阵元之和都为1的矩阵（请读者想一想，这一特点的“物理意义”是什么？）^[3]。而我们的矩阵 \mathbf{H} 却可能有一些列是零向量，从而矩阵元之和为0，它们对应于那些没有对外链接的网页，即所谓的“悬挂网页”（dangling page）^[4]。

上述公式的求解是简单得不能再简单的事情，即

$$\boldsymbol{p}_n = \boldsymbol{H}^n \boldsymbol{p}_0$$

其中 \boldsymbol{p}_0 为虚拟读者初次浏览时访问各网页的几率分布（在佩奇和布林的原始论文中，这一几率分布被假定为是均匀分布）。

三、问题及解决

如前所述，佩奇和布林是用虚拟用户在经过很长——理论上为无穷长——时间的漫游后访问各网页的几率分布，即 $\lim_{n \rightarrow \infty} \mathbf{p}_n$ ，来确定网页排序的。这个定义要想管用，显然要解决三个问题：

(1) 极限 $\lim_{n \rightarrow \infty} \mathbf{p}_n$ 是否存在？

(2) 如果极限存在，它是否与 \mathbf{p}_0 的选取无关？

(3) 如果极限存在，并且与 \mathbf{p}_0 的选取无关，它作为网页排序的依据是否真的合理？

如果这三个问题的答案都是肯定的，那么网页排序问题就算解决了。反之，哪怕只有一个问题的答案是否定的，网页排序问题也就不能算是得到了满意解决。那么实际答案如何呢？很遗憾，是后一种，而且是其中最糟糕的情形，即三个问题的答案全都是否定的。这可以由一些简单的例子看出。比方说，在只包含两个相互链接网页的迷你型互联网上，如果 $\mathbf{p}_0 = (1, 0)^T$ ，极限就不存在（因为几率分布将在 $(1, 0)^T$ 和 $(0, 1)^T$ 之间无穷振荡）。而存在几个互不连通（即互不链接）区域的互联网则会使极限——即便存在——与 \mathbf{p}_0 的选取有关（因为把 \mathbf{p}_0 选在不同区域内显然会导致不同极限）。至于极限存在，并且与 \mathbf{p}_0 的选取无关时它作为网页排序的依据是否真的合理的问题，虽然不是数学问题，答案却也是否定的，因为任何一个“悬挂网页”都能像黑洞一样，把其他网页的几率“吸收”到自己身上（因为虚拟用户一旦进入那样的网

页，就会由于没有对外链接而永远停留在那里），这显然是不合理的。这种不合理效应是如此显著，以至于在一个连通性良好的互联网上，哪怕只有一个“悬挂网页”，也足以使整个互联网的网页排序失效，可谓是“一粒老鼠屎坏了一锅粥”。

为了解决这些问题，佩奇和布林对虚拟用户的行为进行了修正。首先，他们意识到无论真实用户还是虚拟用户，当他们访问到“悬挂网页”时，都不应该也不会“在一棵树上吊死”，而是会自行访问其他网页。对于真实用户来说，自行访问的网页显然与个人的兴趣有关，但对于在平均意义上代表真实用户的虚拟用户来说，佩奇和布林假定它将会在整个互联网上随机选取一个网页进行访问。用数学语言来说，这相当于是把 \mathbf{H} 的列向量中所有的零向量都换成 \mathbf{e}/N （其中 \mathbf{e} 是所有分量都为1的列向量， N 为互联网上的网页总数）。如果我们引进一个描述“悬挂网页”的指标向量（indicator vector） \mathbf{a} ，它的第 i 个分量的取值视 W_i 是否为“悬挂网页”而定——如果是“悬挂网页”，取值为1，否则为0——并用 \mathbf{S} 表示修正后的矩阵，则

$$\mathbf{S} = \mathbf{H} + \frac{\mathbf{e}\mathbf{a}^T}{N}$$

显然，这样定义的 \mathbf{S} 矩阵的每一列的矩阵元之和都是1，从而是一个不折不扣的随机矩阵。这一修正因此而被称为随机性修正（stochasticity adjustment）。这一修正相当于剔除了“悬挂网页”，从而可以给上述第三个问题带来肯定回答（当然，这一回答没有绝对标准，可以不断改进）。不过，这一修正解决不了前两个问题。为了解决那两个问题，佩奇和布林引进了第二个修正。他们假定，虚拟用户虽然是虚拟的，但多少也有一些“性格”，不会完全受当前网页所限，死板地只访问其所提供的链接。具体地说，他们假定虚拟用户在每一步都有一个小于1的几率 α

访问当前网页所提供的链接，同时却也有一个几率 $1-\alpha$ 不受那些链接所限，随机访问互联网上的任何一个网站。用数学语言来说（请读者自行证明），这相当于是把上述 \mathbf{S} 矩阵变成了一个新的矩阵 \mathbf{G} ：

$$\mathbf{G} = \alpha \mathbf{S} + \frac{(1-\alpha)\mathbf{e}\mathbf{e}^T}{N}$$

这个矩阵不仅是一个随机矩阵，而且由于第二项的加盟，它有了一个新的特点，即所有矩阵元都为正，（请读者想一想，这一特点的“物理意义”是什么？）这样的矩阵是所谓的素矩阵（primitive matrix）^[5]。这一修正因此而被称为素性修正（primitivity adjustment）。

经过这两类修正，网页排序的计算方法就变成了

$$\mathbf{p}_n = \mathbf{G}^n \mathbf{p}_0$$

这个算法能给上述问题提供肯定答案吗？是的，它能。因为随机过程理论中有一个所谓的马尔可夫链基本定理（fundamental theorem of Markov chains），它表明在一个马尔可夫过程中，如果转移矩阵是素矩阵，那么上述前两个问题的答案就是肯定的。而随机性修正已经解决了上述第三个问题，因此所有问题就都解决了。如果我们用 \mathbf{p} 表示 \mathbf{p}_n 的极限^[6]，则 \mathbf{p} 给出的就是整个互联网的网页排序——它的每一个分量就是相应网页的访问几率，几率越大，排序就越靠前。

这样，佩奇和布林就找到了一个不仅含义合理，而且数学上严谨的网页排序算法，他们把这个算法称为PageRank，不过要注意的是，虽然这个名称的直译恰好是“网页排序”，但它实际上指的是“佩奇排序”，因为其中的“Page”不是指网页，而是佩奇的名字。这个算法就是谷歌排序

的数学基础，而其中的矩阵 \mathbf{G} 则被称为谷歌矩阵（Google matrix）。

细心的读者可能注意到了，我们还遗漏了一样东西，那就是谷歌矩阵中描述虚拟用户“性格”的那个 α 参数。那个参数的数值是多少呢？从理论上讲，它应该来自于对真实用户平均行为的分析，不过实际上另有一个因素对它的选取产生了很大影响，那就是 $\mathbf{G}^n \mathbf{p}_0$ 收敛于 \mathbf{p} 的快慢程度。由于 \mathbf{G} 是一个 $N \times N$ 矩阵，而 N 为互联网上——确切地说是被谷歌所收录的——网页的总数，在谷歌成立之初为几千万，目前为几百亿（并且还在持续增加），是一个极其巨大的数字。因此 \mathbf{G} 是一个超大型矩阵，甚至很可能是人类有史以来处理过的最庞大的矩阵。对于这样的矩阵， $\mathbf{G}^n \mathbf{p}_0$ 收敛速度的快慢是关系到算法是否实用的重要因素，而这个因素恰恰与 α 有关。可以证明， α 越小， $\mathbf{G}^n \mathbf{p}_0$ 的收敛速度就越快。但 α 也不能太小，因为太小的话，“佩奇排序”中最精华的部分，即以网页间的彼此链接为基础的排序思路就被弱化了（因为这部分的贡献正比于 α ），这显然是得不偿失的。因此，在 α 的选取上有很多折中的考虑要做，佩奇和布林最终选择的数值是 $\alpha=0.85$ 。

以上就是谷歌背后最重要的数学奥秘。与以往那种凭借关键词出现次数所作的排序不同，这种由所有网页的相互链接所确定的排序是不那么容易做假的，因为作假者再是把自己的网页吹得天花乱坠，如果没有真正吸引人的内容，别人不链接它，一切就还是枉然^[7]。而且“佩奇排序”还有一个重要特点，那就是它只与互联网的结构有关，而与用户具体搜索的东西无关。这意味着排序计算可以单独进行，而无需在用户键入搜索指令后才临时进行。谷歌搜索的速度之所以快捷，在很大程度上得益于此。

四、结语

在本文的最后，我们顺便介绍一点谷歌公司的历史。佩奇和布林对谷歌算法的研究由于需要收集和分析大量网页间的相互链接，从而离不开硬件支持。为此，早在研究阶段，他们就四处奔走，为自己的研究筹集资金和硬件。1998年9月，他们为自己的试验系统注册了公司——即如今大名鼎鼎的谷歌公司。但这些行为虽然近乎于创业，他们两人当时却并无长期从商的兴趣。1999年，当他们觉得打理公司干扰了自己的研究时，甚至萌生了卖掉公司的想法。

他们的开价是100万美元。

与谷歌在短短几年之后的惊人身价相比，那简直就是“跳楼大甩卖”。可惜当时却无人识货。佩奇和布林在硅谷“叫卖”了一圈，连一个买家都没找到。被他们找过的公司包括了当时搜索业巨头之一的Excite（该公司后来想必连肠子都悔青了）。为了不让自己的心血荒废，佩奇和布林只得将公司继续办了下去，一直办到今天，这就是谷歌的“发家史”。

谷歌成立之初跟其他一些“发迹于地下室”（one-man-in-basement）的IT公司一样寒酸：雇员只有一位（两位老板不算），工作场所则是一位朋友的车库。但它出类拔萃的排序算法很快为它赢得了声誉。公司成立仅仅3个月，*PC Magazine*杂志就把谷歌列为了年度最佳搜索引擎。2001年，佩奇为“佩奇排序”申请到了专利，专利的发明人为佩奇，拥有者则是他和布林的母校斯坦福大学。2004年8月，谷歌成为了一家初始市值约17亿美元的上市公司。不仅公司高管在一夜间成为了亿万富翁，就连当初给过他们几十美元“赞助费”的某些同事和朋友也得到了足够终

身养老所用的股票回报。作为公司摇篮的斯坦福大学则因拥有“佩奇排序”的专利而获得了180万股谷歌股票。2005年12月，斯坦福大学通过卖掉那些股票获得了3.36亿美元的巨额收益，成为美国高校因支持技术研发而获得的有史以来最巨额的收益之一^[8]。



谷歌公司创始人佩奇（左）和布林（右）

谷歌在短短数年间就横扫整个互联网，成为搜索引擎业的新一代霸主，佩奇和布林的那个排序算法无疑居功至伟，可以说，是数学成就了谷歌^[9]。当然，这么多年过去了，谷歌作为IT界研发能力最强的公司之一，它的网页排序方法早已有了巨大的改进，由当年单纯依靠“佩奇排序”演变为了由200多种来自不同渠道的信息——其中包括与网页访问量有关的统计数据——综合而成的更加可靠的方法。而当年曾给佩奇和布林带来过启示的学术界，则反过来从谷歌的成功中借鉴了经验，如今一些学术机构对论文影响因子（*impact factor*）的计算已采用了类似“佩奇排序”的算法。谷歌的发展极好地印证了培根（Francis Bacon）的一句名言：知识就是力量。

参考文献

- [1] Austin D. How Google finds your needle in the Web's haystack[OL].
<http://www.ams.org/samplings/feature-column/fcarc-pagerank>.
- [2] Battelle J. The birth of Google[J]. Wired, August 2005.
- [3] Brin S, Page L. The anatomy of a large-scale hypertextual web search engine[C]. Seventh International World-Wide Web Conference, Brisbane, Australia, April 14-18, 1998.
- [4] Ibe O. Markov processes for stochastic modeling[M]. Amsterdam: Elsevier Academic Press, 2009.
- [5] Langville A N, Meyer C D. Google's page rank and beyond: the Science of search engine rankings[M]. Princeton: Princeton University Press, 2006.
- [6] Rousseau C, Saint-Aubin Y. Mathematics and technology[M]. Berlin: Springer, 2008.

2010年12月4日写于纽约

[1]本文曾发表于《数学文化》2011年2月刊（山东大学与香港浸会大学合办）。

[2]马尔可夫过程，也称为马尔可夫链（Markov chain），是一类离散随机过程，它的最大特点是每一步的转移概率分布都只与前一步有关。而平稳马尔可夫过程则是指转移概率分布与步数无关的马尔可夫过程（体现在我们的例子中，即 H 与 n 无关）。另外要说明的是，本文在表述上不同于佩奇和布林的原始论文，后者并未使用诸如“马尔可夫过程”或“马尔可夫链”那样的术

语，也并未直接运用这一领域内的数学定理。

[3]在更细致的分类中，这种每一列的矩阵元之和都为1的随机矩阵称为左随机矩阵（left stochastic matrix），以区别于每一行的矩阵元之和都等于1的所谓右随机矩阵（right stochastic matrix）。这两者在应用上基本是等价的，区别往往只在于约定。

[4]这种几乎满足随机矩阵条件，但有些列（或行）的矩阵元之和小于1的矩阵也有一个名称，叫做亚随机矩阵（substochastic matrix）。

[5]确切地说，这种所有矩阵元都为正的矩阵不仅是素矩阵，而且还是所谓的正矩阵（positive matrix）。这两者的区别是：正矩阵要求所有矩阵元都为正，而素矩阵只要求自己的某个正整数次幂为正矩阵。

[6]读者们想必看出来了， \mathbf{p} 其实是矩阵 \mathbf{G} 的本征值为1的本征向量，而利用虚拟用户确定网页排序的思路其实是在用迭代法解决上述本征值问题。在数学上可以证明，上述本征向量是唯一的，而且 \mathbf{G} 的其他本征值 λ 全都满足 $\lambda < 1$ （更准确地说，是 $|\lambda| \leq \alpha$ ——这也正是下文即将提到的 $\mathbf{G}^n \mathbf{p}_0$ 的收敛速度与 α 有关的原因）。

[7]当然，这绝不意味着在网页排序上已不可能再做假。相反，这种做假在互联网上依然比比皆是，比如许多广告或垃圾网页制造者用自动程序到各大论坛发帖，建立对自己网页的链接，以提高排序，就是一种常见的做假手法。为了遏制做假，谷歌采取了很多技术手段，并对有些做假网站采取了严厉的惩罚措施。这种惩罚（有时是误罚）对于某些靠互联网吃饭的公司有毁灭性的打击力。

[8]从投资角度讲，斯坦福大学显然是过早卖掉了股票，否则获利将更为丰厚。不过，这正是美国名校的一个可贵之处，它们虽擅长从支持技术研发中获利，却并不唯利是图。它们有自己的原则，那就是不能让商业利益干扰学术研究。为此，它们通常不愿长时间持有特定公司的股票，以免在无形中干扰与该公司存在竞争关系的学术研究的开展。

[9]有些读者对“是数学成就了谷歌”这一说法不以为然，认为是佩奇和布林的商业才能，或将数

学与商业结合起来的才能成就了谷歌。这是一个见仁见智的问题，看法不同不足为奇。我之所以认为是数学成就了谷歌，是因为谷歌当年胜过其他搜索引擎的地方只有算法。除算法外，佩奇和布林当年并无其他胜过竞争对手的手段，包括商业手段。如果让他们去当其他几家搜索引擎公司的老总，用那几家公司的算法，他们是不可能脱颖而出的；而反过来，如果让其他几家搜索引擎公司的老总来管理谷歌，用谷歌的算法，我相信谷歌依然能超越对手。因此，虽然谷歌后来确实用过不少出色的商业手段（任何一家那样巨型的公司都必然有商业手段上的成功之处），而当年那个算法在今天的谷歌——如正文所述——则早已被更复杂的算法所取代，但我认为谷歌制胜的根基和根源在于那个算法，而非商业手段，因此我说“是数学成就了谷歌”。

第二部分 物理



绘画：张京

从巴西的蝴蝶到得克萨斯的飓风^[1]

一、决定论

在本书《时间旅行：科学还是幻想？》一文的第四节中，我们曾提到混沌理论中的一个概念：蝴蝶效应（butterfly effect）。这个效应也被称为对初始条件的敏感依赖性，指的是在某些——通常是非线性的——物理体系中，初始条件的细微改变有可能对体系的未来演化产生巨大影响。它的一种很富诗意的形容，是说巴西的一只蝴蝶拍动翅膀产生的空气扰动，有可能演变成美国得克萨斯州的一场飓风。这也是蝴蝶效应这一名称的主要由来。本文将对这一概念及其历史做一个简单介绍。

我们知道，人类描述自然的努力，很大程度上体现在对自然现象的时间演化进行描述上。这种描述在许多方面都取得了很大的成功。早在300多年前，英国科学家牛顿（Isaac Newton）就建立了我们称为牛顿力学的理论体系，对小至钟摆、陀螺，大至行星运动的各种自然现象的时间演化做出了极为精确的描述。1846年，天文学家们在牛顿力学所预言的位置近旁发现了几十亿千米之外的太阳系第8大行星——海王星，成为牛顿力学最辉煌的成就之一^[2]。

牛顿力学的成功，除了体现在对某些自然现象时间演化的极为精确的描述外，还留下了一个非常重要的遗产，那就是决定论的思想。按照这一思想，从一个物理体系在某一时刻的状态，可以推算出它在任何其

他时刻的状态。人们后来知道，牛顿力学本身只适用于描述一定范围内的力学现象，但它所留下的决定论思想却适用于几乎所有已知的物理定律，甚至在一定程度上包括了被公认为是非决定论性的量子力学^[3]。

那么，决定论思想所具有的如此广泛的适用性，是否意味着我们在原则上可以对物理现象作出精确预言呢？在很长一段时间里，答案被认为是肯定的。但是，与这种被认为原则上可以做到的精确预言形成鲜明对比的，是实际上能精确求解的物理问题的稀少。以天体运动为例，人们能精确求解的只有二体问题。一旦把太阳、地球和月球这三个最熟悉的天体同时考虑进去，就没法精确求解了^[4]。又比如流体运动，能精确求解的只有一些非常理想的情形，一旦把像黏滞性那样最常见的现实性质考虑进去，也就没法精确求解了。物理学家们能精确求解的问题，大都附加了各种简化条件。而真正的自然现象几乎从来都不满足那些条件，从而几乎没有一个是能精确求解的。

幸运的是，在那些无法精确求解的问题中，有一部分非常接近于某些能精确求解的问题。比如地球绕太阳的运转，所有其他天体的影响都相当微小，因此这一问题非常接近于能精确求解的二体问题。而且这两者的差异还可以通过各种手段加以弥补。正是由于这些近似手段——包括数值近似——的存在，使得物理学家们虽然很少能精确求解问题，却依然能对很多自然现象的演化做出非常成功的描述。

二、早期研究

但是，任何近似手段都必然有误差，因此近似手段的有效性有赖于对误差的控制。随着研究的深入，物理学家们开始遇到了一些无法用近似手段来有效处理的问题。这些问题中有许多都具有蝴蝶效应，它使误差变得不可控制。19世纪末，法国科学家庞加莱（Henri Poincaré）在对三体问题的研究中发现了一些这样的问题。他在《科学与方法》一书中写道：“初始条件的微小差异有可能在最终的现象中导致巨大差异”，“预言变得不可能”。这或许是对蝴蝶效应最早的明确描述^[5]。除三体问题外，流体力学中的湍流问题也是一种无法用近似手段来有效处理的问题。据说德国物理学家海森伯（Werner Heisenberg）曾经表示，有机会向上帝提问的话，他想问上帝为什么会有相对论？以及为什么会有湍流？他并且补充说：“我确信上帝知道第一个问题的答案。”——言下之意是上帝也未必知道为什么会有湍流。

当科学家们接触到包含蝴蝶效应的问题或现象时，科幻小说家们也在用自己独特的方式描述着类似的现象。比如1955年，美国科幻小说家阿西莫夫（Isaac Asimov）写了一部小说，叫做《永恒的终结》（*The End of Eternity*）。在这部小说中，阿西莫夫描述了一群生活在物理时间之外的人，他们可以对人类历史进行修正，使其更加完美。但他们企图为人类创造一个完美历史的努力，在无形中扼杀了人类的创造与探索能力，使其在与外星生命的竞争中一败涂地。幸运的是，人类后来发现了这一点，并通过时间旅行的手段挽回了一切。在这部小说中阿西莫夫提到：对历史的每一次微小改变，都有可能以一种无法精确预言的方式改变数百万人的人生轨迹，这与蝴蝶效应的表述显然有着极大的相似性。这种出现在科幻小说中的近乎先知先觉的描述，初看起来很令人吃惊，

其实并不奇怪。因为现实世界本身就是一种最复杂的自然现象，像蝴蝶效应那样的东西，远在它成为科学研究的对象之前，就早已出现在了人们的日常经验中。所谓“差之毫厘，谬之千里”、“牵一发而动全身”等，都在一定程度上体现了这种效应。

但从那些日常体验上升为明确的理论表述，则是一个困难得多的问题。

从19世纪末到20世纪中叶，经过庞加莱、利雅普诺夫（Aleksandr Lyapunov）、弗兰克林（Philip Franklin）、马科夫（Andrei Andreevich Markov）、伯克霍夫（George David Birkhoff）等人的一系列研究，人们对这个困难得多的问题终于有了一定的认识。人们发现，对于满足一定条件的物理体系来说，只有周期性或近周期性（near periodic）的运动才不会因为初始条件的细微改变而产生剧烈变动。依照这个结果，如果运动是非周期性的，那么初始条件的细微改变就会对体系的演化造成巨大影响。因此，这个结果不仅确立了蝴蝶效应的存在，而且还对它的产生条件给出了一定的描述。但是，那时候人们最感兴趣的只是周期运动，因此有关非周期运动的结果虽可作为推论得到，在当时的学术文献中却极少提及。正因为如此，十几年后当美国科学家洛伦兹（Edward Norton Lorenz）在数值计算中再次遭遇蝴蝶效应的时候，依然感到了极大的惊讶。也正因为如此，发现蝴蝶效应的荣誉在很大程度上被后人归到了洛伦兹的头上。

三、模拟天气

洛伦兹是一位资深的气象学家，早在“二战”时期就在美国军方机构从事气象预测研究。战争结束后，洛伦兹来到了麻省理工学院

（MIT），继续从事研究工作。从理论上预测气象变化——尤其是给出长期预测——是气象学家们梦寐以求的目标。但这一目标的实现却始终困难重重。这种困难是不难理解的，因为地球的大气层是一个巨大的流体系统，所有流体力学所具有的复杂性，包括那个连上帝也未必知道起源的湍流问题，都会出现在大气层中。更何况，大气层的行为与海洋、地表、日照等各种复杂的外部条件都有着密切关系；而且大气层的组成相当复杂，其中有些组成部分——如水汽——的形态还常常会在气态、液态及固态之间变化。所有这一切，都使得气象预测成为了一个极其困难的课题。

在洛伦兹从事气象研究的时候，从理论上预测气象变化主要有两类方法。一类被称为动力气象学（dynamic meteorology），这类方法主要是把大气层看作一个流体系统，然后选取一些重要的物理量——如温度、风速等——进行研究。由于问题的高度复杂，人们还把大气层像切蛋糕一样分割成许多区域，每个区域都用一个点来代表。显然，这是极其粗糙的近似，但即便如此，整个大气层的状态往往还是需要几百万甚至更大数目的变量来描述^[6]。换句话说，即便是求解一个非常粗糙的气候模型，往往也需要处理带有几百万个未知数的方程组。这无疑是极其困难的（但不是完全没有希望的）。除了动力气象学外，还有一类方法被称为天气学（synoptic meteorology），这类方法的特点是把对气候影响最大的一些大气结构，比如各种气旋，直接作为研究对象。天气学所使用的规律，有许多是描述那些大气结构的经验规律，而不是像流体力

学那样系统性的物理理论。从这个意义上讲，天气学不如动力气象学那样基本。但天气学的优点，是把从动力气象学角度看非常复杂的某些大气结构作为了基本单元，从而有着独特的简化性。

洛伦兹所采用的主要是天气学方法。经过大量的简化后，洛伦兹得到了一个含有14个变量，且其中有一到两个变量的影响可以忽略的模型。但即使那样的模型用手工计算也是非常困难的，于是洛伦兹决定借助计算机的帮助。当时是1959年，距离个人计算机的问世还有二十几年。洛伦兹所使用的机器用今天的标准来衡量是极为简陋的：体积庞大，噪声惊人，内存却只有今天普通个人计算机内存的百万分之一。经过几个月的努力（主要是编程），洛伦兹终于在那台机器上运行起了他的模拟天气。

四、奇怪的结果

日子平静地流逝着，洛伦兹与同事们间或地就模拟天气的演变打上一些小赌，聊以消遣。终于有一天，洛伦兹决定对某一部分计算进行更为仔细的分析。于是他从原先输出的计算结果中选出了一行数据——相当于某一天的天气状况——作为初始条件输入了程序。机器从那一天的数据开始了运行，洛伦兹则离开了办公室，去喝一杯悠闲的咖啡。中国的神话故事中有所谓“洞中方一日，世上已千年”的传说，洛伦兹的那杯咖啡就喝出了那样的境界。一个小时后，当他回到实验室时，他的模拟世界已经运行了两个月。洛伦兹一看结果，不禁吃了一惊！因为新的计算结果与原先的大相径庭。

这一结果为什么令人吃惊呢？因为这次计算所用的初始条件乃是从旧数据中选出来的。既然初始条件是旧的，所得的结果——在与旧数据可以比较的范围内——理应也跟旧数据相同，却怎么会大相径庭呢？洛伦兹的第一个反应是机器坏了——这在当时是经常发生的事情。但是，当他对结果做了更仔细的检验后，很快排除了那种可能性。因为他发现，新旧计算的结果虽然最终大相径庭，但在一开始却很相似，两者的偏差是在经过了一段指数增长过程之后才彻底破坏相似性的。如果机器坏了，是没有理由出现这种“有规律”的偏差的。

既然机器没有问题，那么究竟是什么原因造成了新旧计算之间的巨大偏差呢？洛伦兹很快找到了答案。原来，洛伦兹的程序在运行时保留了十几位有效数字，但在输出时为了让所有变量的数值能打印在同一行里，他对每个变量都只保留了小数点后3位有效数字。因此，当洛伦兹把以前输出的数据——即所谓旧数据——作为初始条件输入新一轮计算

时，它与原先计算中保留了十几位有效数字的数据相比，已经有了微小的偏差。洛伦兹的计算表明，在他的模拟系统中，这些微小的偏差每隔4天就会翻一番，直至新旧数据间的相似性完全丧失为止。

这正是蝴蝶效应！

由于蝴蝶效应的存在，洛伦兹意识到长期天气预报是注定不可能有高精度的。因为我们永远不可能得到绝对精确的初始条件，而且由于任何计算设备的内存都是有限的，我们在计算过程中也永远不可能保留无限的精度，所有这些误差都会因蝴蝶效应的存在而迅速（指数性地）扩大，从而不仅使一切高精度的长期气象预测成为泡影，而且葬送了建立在决定论思想之上的对物理现象进行精确预言的梦想^[7]。

蝴蝶效应的发现还让洛伦兹回忆起一件他念本科时发生的事情。那是在20世纪30年代，当时他所在的镇上有许多学生迷上了弹球游戏（pinball game），那是一种让小球在一张插有许多小针的倾斜桌子上经过多次碰撞后进入特定小孔的游戏。当地政府曾想以禁止赌博为由禁止这种游戏，但游戏的支持者们争辩说这不是赌博，而是一种有关击球准确度的技巧比赛。他们的理由一度说服了政府官员，因为当时大家并不知道弹球游戏其实包含了蝴蝶效应，从而无论多高明的技巧都是无济于事的。

五、从蝴蝶到飓风

发现蝴蝶效应后的第二年，即1960年，洛伦兹在一次学术会议上粗略地提及了自己的发现，但没有发表详细结果。会议之后，洛伦兹感到自己的模型仍然太复杂，他决定寻找更简单的模型。1961年，他从同事索兹曼（Barry Saltzman）那里得到了一个只含7个变量（即比他自己的模型少了一半的变量）的流体力学模型^[8]。经过研究，洛伦兹很快发现，在索兹曼的模型中，有4个变量的数值很快就会变得可以忽略。因此，这一模型的真实行为可以用一个只含3个变量的方程组来描述，这个只含3个变量的方程组后来被冠上了洛伦兹的大名，称为洛伦兹方程组（Lorenz equations）。利用这一方程组，洛伦兹再次确认了蝴蝶效应的存在^[9]，并于1963年在《大气科学杂志》（*Journal of the Atmospheric Sciences*）上发表了题为《确定性非周期流》（*Deterministic Nonperiodic Flow*）的论文，正式公布了自己的结果。

不过，无论是洛伦兹的原始论文，还是此后若干年内的其他有关著作，都没有直接使用“蝴蝶效应”这一名称。洛伦兹本人有时用海鸥造成的大气扰动来比喻初始条件的细微改变。“蝴蝶”这一“术语”的使用是在9年后的1972年。那一年洛伦兹要在华盛顿的一个学术会议上做报告，却没有及时提供报告的标题。于是会议组织者梅里利斯（Philip Merilees）“擅作主张”地替洛伦兹拟了一个题目：《巴西的蝴蝶拍动翅膀会引发德克萨斯的飓风吗？》（*Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?*）。就这样，美丽的蝴蝶随着梅里利斯的想象飞进了科学术语之中^[10]。

除上述原因之外，“蝴蝶效应”的得名还有另外一个原因，那就是洛

洛伦兹模型中有一个所谓的奇怪吸引子（strange attractor），它的形状从一定的角度看很像一只展翅的蝴蝶（图1）。不过“蝴蝶效应”这一名称的最终风行，在很大程度上要归因于美国科普作家格雷克（James Gleick）的科普作品《混沌：开创新科学》（*Chaos: Making a New Science*）。这部被译成了多国文字，对混沌理论（蝴蝶效应是混沌理论的一部分）在世界范围内的热播起了极大促进作用的作品的第一章的标题就是《蝴蝶效应》。2004年，蝴蝶效应甚至被搬上了银幕，成为一部科幻影片——虽然是不太成功的影片——的片名。

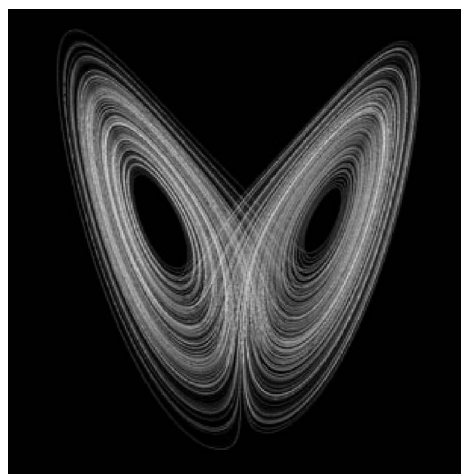


图1 洛伦兹奇怪吸引子

蝴蝶效应及混沌理论在世界范围内的风行，一度使许多人产生一种错觉，以为物理学的又一次革命到来了。在这种“激情”的鼓舞下，这一领域涌现出了大量文章，其中包括不少低水平及浮夸的工作。从物理学的角度讲，蝴蝶效应及混沌理论并不包含新的原理性的东西，它们对物理学的最大启示是：形式上简单的物理学定律有可能包含巨大的复杂性，从而有可能解释比我们曾经以为的更为广阔的自然现象。这一点早在洛伦兹的论文发表之前，就已经被一些物理学家注意到了。20世纪60年代初，美国物理学家费恩曼（Richard Feynman）在给本科生讲课——那些课程的内容后来汇集成了著名的《费恩曼物理学讲义》（*The Feynman Lectures on Physics*）时，就非常清晰地阐述了这一点。他在介绍了流体力学中的若干复杂性之后这样写道：

对物理学怀有莫名恐惧的人常常会说，你无法写下一个关于生命的方程式。嗯，也许我们能够。事实上，当我们写下量子力学方程式 $H\Psi = i\partial\Psi/\partial t$ 的时候，我们很可能就已在足够近似的意义上拥有了这样的方程式。我们刚才就看到了事物的复杂性可以多么容易且富有戏剧性地逃脱描述它们的方程式的简单性。

费恩曼曾经希望人类的下一次智力启蒙会带给我们理解物理定律复杂内涵的方法。混沌理论的发展部分地体现了费恩曼的希望，但今天我们对这一领域的了解，在很大程度上依赖于计算技术的发展，与真正的智力启蒙还有一定的距离。真正的智力启蒙究竟会出现在什么时候？也许就像洛伦兹的天气一样，谁也无法准确预测，但我们会拭目以待。

2006年7月23日写于纽约

2014年9月24日最新修订

[1]本文的一个缩略修改版曾发表于《科幻世界》2007年第1期（科幻世界出版社出版）。

[2]不过后来的研究表明，海王星在距离理论预言非常近——相差不到 1° ——的位置上被发现有一定的偶然性。关于这一点，可参阅拙作《那颗星星不在星图上：寻找太阳系的疆界》的第20章。

[3]量子力学的状态演化是决定论性的，但量子测量过程是否也是决定论性的，则有一定的争议（虽然非决定论性的观点明显占优）。

[4]这还是在假定引力是由牛顿万有引力定律所描述的情况下，如果改用广义相对论，则连二体问题也无法严格求解。

[5]不过《科学与方法》是一部科学哲学著作，庞加莱在自己的学术论文中并未明确表述过类似的结论。

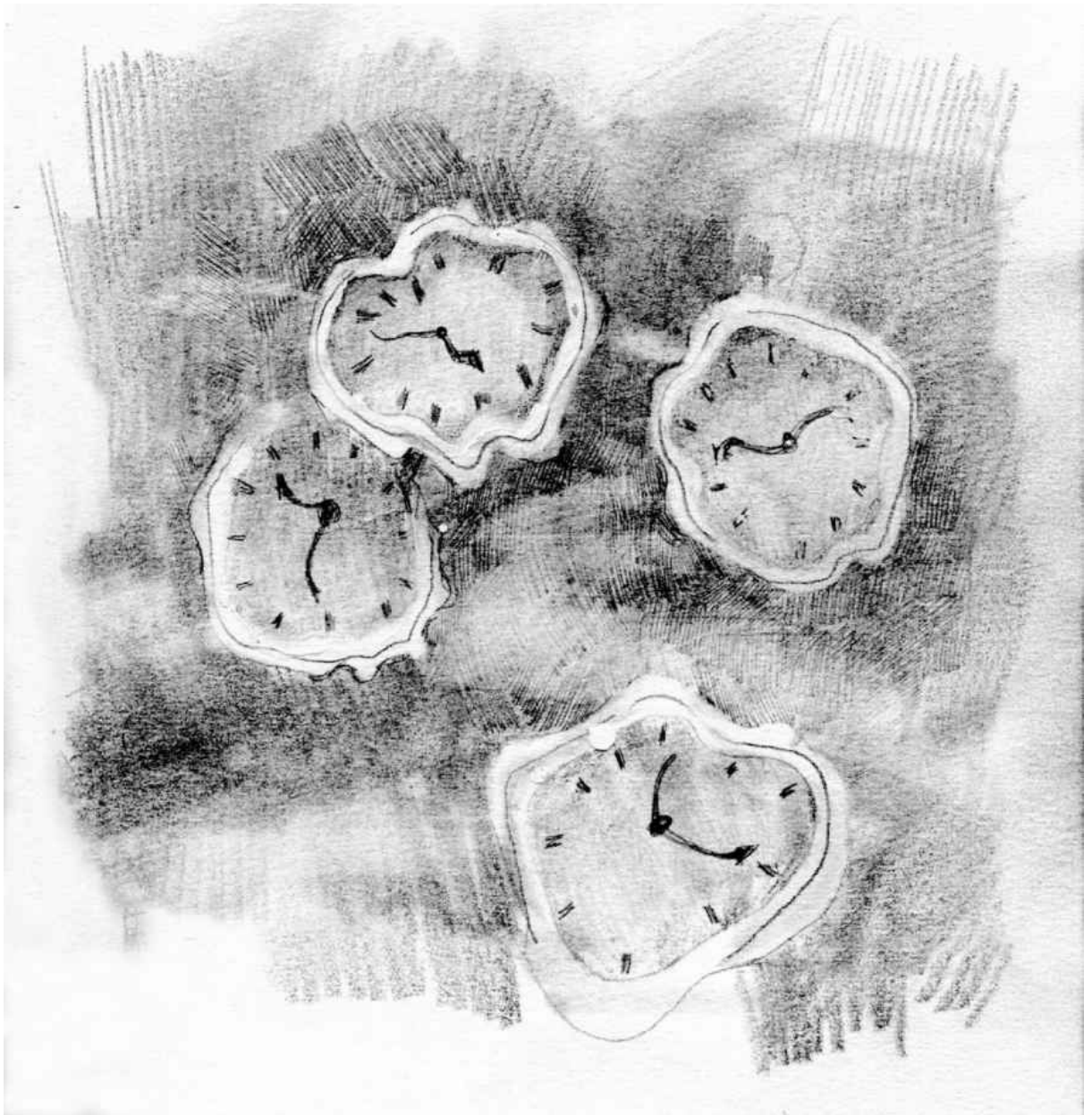
[6]举个例子来说，如果把大气层用长、宽、高分别为100千米、100千米及100米的单元进行分割，则描述整个大气层——假定高度为30千米——的温度与风速所需的变量总数大约为500万。分割越细、引进的物理量越多，所需的变量数目也就越大。

[7]严格地讲，由于无法得到精确的初始条件，以及无法在计算过程中保留无限的精度，即便没有蝴蝶效应，绝对精确的预言也是不可能的。但在没有蝴蝶效应的情况下，误差的影响往往是可控制的，蝴蝶效应的出现使误差的影响变得不可控制。另外需要说明的是，这里所说的“葬送了建立在决定论思想之上的对物理现象进行精确预言的梦想”与建立在微分方程解的存在及唯一性基础之上的决定论本身不是一回事，后者不会因为蝴蝶效应而破灭。

[8]索兹曼与20世纪上半叶的那些科学家一样，对周期运动更感兴趣，因此没能在自己的模型上做出像洛伦兹那样的发现，虽然他在自己的模型中也已经发现了一些非周期性的解。

[9]在这一点上，洛伦兹很受幸运女神的眷顾。他的方程组中含有一个被称为普朗特常数（Prandtl constant）的参数，这个参数对于水大约为10，对于空气则大约为1。洛伦兹与索兹曼都是气象学家，他们采用的数值原本应该是对应于空气的1，但实际上两人却都采用了对应于水的10。后来的研究发现，如果当时他们采用了对应于空气的普朗特常数，那个模型的解将是周期性的，洛伦兹将不可能得到他所需要的结果。

[10]不过那篇演讲的全文当时并未发表。另外需要提醒读者的是：蝴蝶效应的这一通俗表述有一定的误导性，容易让人以为在“蝴蝶拍动翅膀”与“得克萨斯的飓风”之间存在直接的因果联系。事实上，“蝴蝶拍动翅膀”和“得克萨斯的飓风”只是泛指初始条件的细微改变和体系未来演化的巨大变化，“得克萨斯的飓风”的物理起因有赖于无数的因素，绝非只是“蝴蝶拍动翅膀”。



绘画：张京

关于时钟佯谬

一、时钟佯谬简史

在相对论的历史上，曾出现过一些流传很广的佯谬——也可以说是意外。之所以说是意外，是因为一些知名物理学家也参与了某些话题的讨论，给出的答案却不尽相同，从而使被讨论的话题变得更像佯谬。时钟佯谬（clock paradox）就是其中最著名的一个。

时钟佯谬源于一个很简单的问题：在惯性参照系中有两个彼此校准了的时钟，一个保持静止，另一个沿闭合路线运动后回到原地，问两个时钟重新相遇时哪个时钟慢了？

最早对这个问题作出回答的当然不是别人，而是爱因斯坦（Albert Einstein）本人。他在狭义相对论的开山之作《论动体的电动力学》（*On the Electrodynamics of Moving Bodies*）中对这一问题给予了明确回答，答案是运动时钟慢了，理由是狭义相对论的时钟延缓（time retardation）效应^[1]。但不知是没往那个角度考虑，还是觉得那不是问题，爱因斯坦只在静止时钟参照系中回答了这一问题，而未如许多后人所做的那样，将静止时钟参照系与运动时钟参照系视为对等来进行分析，从而没有触及后来被称为“佯谬”的东西——即两个时钟均认为自己静止，对方运动，从而在表观上彼此矛盾地均认为是对方慢了。

文献中最早触及时钟佯谬，并试图给予解释的知名人士是法国物理

学家朗之万（Paul Langevin）。他在1911年发表的一篇题为《空间和时间的演进》（*The Evolution of Space and Time*）的文章中提出：解释时钟佯谬的要点，是注意到运动时钟经历了加速运动，而静止时钟没有经历加速运动。在那篇文章中，他还引入了两条很受后人欢迎的“人性化措施”：一是将时钟换成人，二是采用两人互发光信号的方法来比较各自经历的时间。朗之万这篇文章影响了很多，时钟佯谬的别名“双生子佯谬”（twin paradox）据说就是他将时钟换成人所引发的（图2）。直至今日，许多初等教材及科普读物仍采用朗之万的方法分析时钟佯谬（有些著作还将两人互发光信号改成更“亲密”地互数对方心跳）。

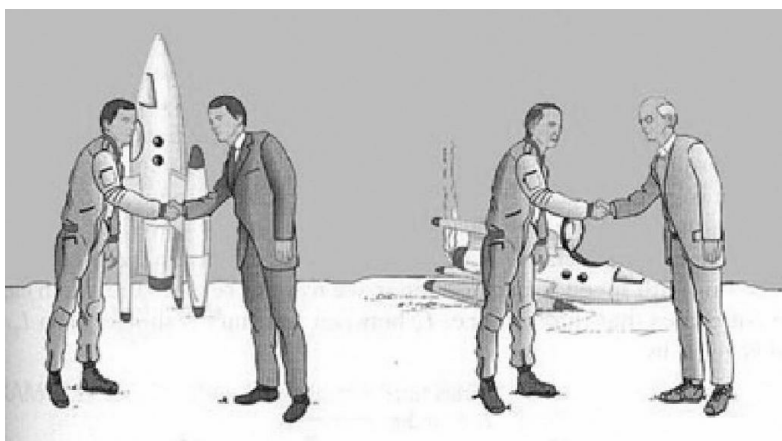


图2 时钟佯谬的别名是双生子佯谬

继朗之万之后，德国物理学家冯·劳厄（Max von Laue）也对时钟佯谬提出了解释。他在1912年发表的一篇题为《两种反相对论意见及对它们的反驳》（*Two Objections against the Theory of Relativity and Their Refutation*）的文章中提出：解释时钟佯谬的要点，是注意到运动时钟在从远离转为返回的过程中更换了参照系，而静止时钟没有更换参照系。他的这一看法也被一些人采纳，成为分析时钟佯谬的切入点之一。

爱因斯坦本人则似乎直到广义相对论发表之后，才对时钟佯谬中

的“佯谬”部分发表看法。他在1918年发表的一篇题为《关于反相对论意见的对话》（*Dialog about Objections against the Theory of Relativity*）的文章中运用自己的“独门武器”等效原理提出了一种看法。他认为解释时钟佯谬的要点，是注意到运动时钟受到了与加速场等效的引力场的影响，而静止时钟没有受到那样的影响。由于引力场中的时钟延缓是广义相对论的推论，加上爱因斯坦在相对论领域的巨大威望，他的解释在很长的时间里被视为了时钟佯谬的正解。20世纪上半叶的很多相对论名著，比如奥地利物理学家泡利（Wolfgang Pauli）的《相对论》（*Theory of Relativity*）、丹麦物理学家莫勒（Christian Møller）的《相对论》（*Theory of Relativity*）、美国物理学家托曼（Richard C. Tolman）的《相对论、热力学及宇宙学》（*Relativity, Thermodynamics, and Cosmology*）等都采用了爱因斯坦的观点，认为时钟佯谬需要用广义相对论来解释。甚至连后来出版的一些知名著作，比如20世纪70年代出版的日本物理学家汤川秀树的《经典物理学》，也认为时钟佯谬“在狭义相对论框架中考虑时是佯谬，但若考虑广义相对论就不再是佯谬了”。

不过，以上这些物理学家对时钟佯谬的解释虽各有各的侧重点，而且从现代观点来看，都不够切中要害，但他们的结论是一致的，并且也是正确的，那就是运动时钟慢了（这一结论后来得到了实验证实）。时钟佯谬作为“佯谬”的正史大体就是这些，但在结束本节之前，有一段“外史”必须提一下，因为时钟佯谬作为“佯谬”的名声，在很大程度上其实是被那段“外史”搅起来的。那段“外史”就是英国哲学家丁格尔（Herbert Dingle）在20世纪50年代末对相对论的猛烈攻击，而那攻击的主要目标就是时钟佯谬。丁格尔在攻击中先是认为两个时钟应显示相同时间，遭驳斥后又转而宣称相对论的预言与经验不符（实际上有关时钟佯谬的预言当时就已得到了 μ 子衰变实验的支持），甚至连数学上显而易见的洛伦兹变换存在逆变换这一基本事实，也被他斥为明显不可能。

这样一个用上海话讲根本就“拎勿清爽”的人物照说是不配在本节中被提及的，但历史有时是充满惊奇的，这位“拎勿清爽”的丁格尔先生在1951—1953年间竟担任过英国皇家天文学会（Royal Astronomical Society）的主席^[2]，并且还是英国科学史学会（British Society for the History of Science）的创始成员之一以及1955—1957年间的主席。也许是因为这些背景的缘故，几十位物理学家对他那破绽百出的文字进行了认真驳斥，从而构成了时钟佯谬历史上一段虽无价值，却引人注目的“外史”，并在最大程度上成就了时钟佯谬作为“佯谬”的名声。

二、时钟佯谬简析

以上就是时钟佯谬的简史，对于我的读者来说，还可以补充一段史上最小的八卦，那就是我“小时候”也曾认同过广义相对论才是时钟佯谬正解的看法，在自己网站（<http://www.changhai.org/>）的昔日版本中还贴过怀旧之作，对某些基于朗之万和冯·劳厄思路的解释进行了“呛声”。也许是因此之故，常有网友问及时钟佯谬。从这个意义上讲，本文可算是一篇还债之作——还那怀旧之作引发的文债。

这篇还债之作之所以拖到今天才写，不是企图“赖债”，而是因为时钟佯谬的现代解释实在太简单了，简直就是“一句话解释”，就算加上注解，似乎也构不成一篇文章。当然，这种估计如今看来显然是错误的，因为真要写的话，几乎没什么话题是构不成文章的。

好了，现在言归正传，谈谈时钟佯谬的现代解释。自20世纪70年代以来，有关相对论的许多教材和专著，比如沃尔德（Robert M. Wald）的《广义相对论》（*General Relativity*）、托雷提（Roberto Torretti）的《相对论与几何》（*Relativity and Geometry*）、伦德勒（Wolfgang Rindler）的《相对论》（*Relativity*）、萨克斯（Rainer Sachs）等人的《数学家用广义相对论》（*General Relativity for Mathematicians*）、米斯纳（Charles W. Misner）等人的《引力》（*Gravitation*）、塞克斯尔（Roman U. Sexl）等人的《相对论、群论、粒子》（*Relativity, Groups, Particles*）等都采用了几何语言来阐述时钟佯谬，这就是所谓时钟佯谬的现代解释。在中文文献中，梁灿彬等人的《微分几何入门与广义相对论》也采用了现代解释，中文读者可以参考^[3]。

那么究竟什么是时钟佯谬的现代解释呢？我没有忽悠诸位，它的要

点确实只有一句话，那就是：时钟记录的是自己的世界线长度。在时钟佯谬中，之所以是运动时钟慢了，原因就是它的世界线长度较短。这里唯一要说明的是，所谓“世界线长度”指的是闵科夫斯基空间中的长度[4]，它与普通空间中的长度有一个最大的区别，那就是前者的测地线（即“直线”）长度是极大值而不是极小值。（请读者想一想，这是闵科夫斯基空间的什么特点造成的？）但无论闵科夫斯基空间还是普通空间，有一点是共同的，那就是长度是坐标变换下的不变量，从而与参照系或坐标系的选择无关[5]。

在这样的几何语言下，时钟佯谬的结论，即运动时钟比静止时钟慢，不过是对两个时钟的世界线长度不同这一简单事实的简单陈述而已，并不比普通空间中两条曲线的长度不同来得奥妙，更没有任何佯谬可言。这一点是如此的显而易见，以至于前面提到的《相对论与几何》一书的作者托雷提感慨道：“相对论时钟是类时世界线上的里程表，假如人们对这一事实有过更多关注，那么在所谓时钟佯谬上付出过的很多努力就可以省掉了。”

但话虽如此，如果我们在这里就结束本文，有些读者也许会感到失望，因为在时钟佯谬的传统讨论中，人们曾花大力气讨论运动时钟参照系，试图说明该参照系也能理解运动时钟变慢这一结论。即便那些努力如今“可以省掉了”，但若不把那最令人困惑的运动时钟参照系单独拿出来，更直接地讨论一下，似乎多少有些偷懒的感觉。为了“抚平”这种感觉，我们再多说几句。

如前所述，在时钟佯谬的现代解释中，运动时钟之所以慢了，原因是它的世界线长度较短。如果画出时空图的话，静止时钟的世界线是直线，运动时钟的世界线是曲线（参阅图3），两者起始点相同，但曲线的长度较短（因为是闵科夫斯基空间）。这一切当然都是几何语言。那

么，在这种语言中运动时钟参照系是什么呢？它就是把运动时钟的世界线视为直线，而把静止时钟的世界线视为曲线的坐标系。这种坐标系其实我们并不陌生，它就是曲线坐标系——把运动时钟的世界线作为时间轴的曲线坐标系。明白了这一点，运动时钟参照系里的问题就迎刃而解了，因为曲线坐标系虽然完全合法，而且确实能在表观上使两条世界线的“曲”、“直”互换，却不会改变它们的长度，从而不会改变时钟佯谬的结论，因为曲线坐标系有一个众所周知的“副作用”，那就是会改变度规的形式，使之

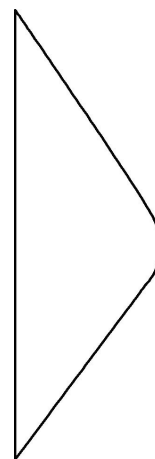


图3 时钟佯

不再是闵科夫斯基度规 $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$ 或欧几里得度规 $ds^2 =$

谬的时空图

$\delta_{ij} dx^i dx^j$ 。比如极坐标下的度规是 $ds^2 = dr^2 + r^2 d\theta^2$ 而不是

$ds^2 = dr^2 + d\theta^2$ 。正是这种度规改变抵消了“曲”、“直”互换的影响，使得长度不变，从而保证了时钟佯谬的结论不变[6]。

当然，这一切其实就是对“长度是坐标变换下的不变量”这一简单事实的繁琐说明，只不过这样一说明，或许显得更像是“解释”而已。另外，它也示范了一种方法，即当我们对时钟佯谬的某个方面感到困惑时，想想它在几何语言下的对应，以及在普通空间中的类比，往往会豁然开朗。

在本节的最后，我们评论一下“时钟佯谬需要用广义相对论来解释”这一流传很广的观点。很明显，时钟佯谬的现代解释并不支持这种观点。时钟佯谬作为闵科夫斯基空间中的现象，是完全可以，并且也应该用狭义相对论来解释的——正如上述现代解释所做的那样。事实上，在闵科夫斯基空间中无论采用什么参照系或坐标系，都不可能使四维曲率张量非零，从而不可能出现曲率意义下的引力场。不仅如此，迄

今为止除上述现代解释外，对时钟佯谬的任何其他解释都是针对特例或近似的。比如朗之万和冯·劳厄的解释通常只被用于运动时钟匀速远离，再匀速飞回的特例；爱因斯坦的解释则往往要采用广义相对论的弱场近似。与之相比，时钟佯谬的现代解释完全不受那些特例或近似的约束，从而有极大的普适性。哪怕两个时钟都作任意复杂的类时运动，现代解释依然适用（传统解释则会变得苦不堪言）。甚至当我们把时钟佯谬的舞台由闵科夫斯基空间搬到更复杂的空间，从而越出狭义相对论的范围时，现代解释依然适用（只需增添一个非平凡的背景度规即可）^[7]。

三、关于理想时钟

在结束本文前，我们还要讨论一个衍生话题：什么是时钟？之所以要讨论这个话题，是因为时钟佯谬的传统解释，尤其是爱因斯坦的思路，很容易产生一个与“什么是时钟？”密切相关的问题，那就是加速度究竟会不会对时钟产生影响？关于这个问题，许多现代教材及专著——比如前面提到的托雷提、塞克斯尔、伦德勒、米斯纳等人的著作——都给出了明确回答，我们在这里作一个简单介绍。

首先要指出的是，对于具体的时钟来说，这个问题的答案显然与时钟的结构有关（而且大都是肯定的），比如对加速场中的摆钟来说，加速度越大，摆动的周期就越短，如果我们用这种摆钟的摆动次数来计时，加速度对它显然是有影响的。又比如对人来说，如果我们将生理节律作为时钟——就像郎之万所做的那样，它显然也会受加速度影响，在足够大的加速度下——对飞行员来说是 $10g$ 以上，对本文作者来说估计 $5g$ 就够了——甚至会“停止计时”（一命呜呼）。不仅宏观世界的时钟如此，曾被用来验证相对论的原子钟，严格讲也是会受加速度影响的，因为它的能级结构与包括加速场在内的各种外场有关。甚至连最早对时钟延缓效应作出实验判决的 μ 子的衰变，我们也并不肯定它不会受加速度影响。只不过，对于微观世界的时钟来说，与它内部的微观相互作用相比，加速场的影响往往是微乎其微的，因此当我们采用微观世界的时钟时，通常都能忽略加速度的影响。

但无论加速度对具体时钟的影响是有还是无，是大还是小，有一点是肯定的，那就是我们并不认为像摆钟受加速度影响，或本文作者在 $5g$ 的加速度下“停止计时”那样的效应反映了时间的固有性质。相反，我们

认为那是具体时钟的缺陷导致的表观效应，是可以、并且必须校正的。我们真正关心的是反映时间本质的时钟，即所谓的理想时钟。本文所说的时钟除非有特别说明，指的也全都是理想时钟。

因此我们的问题其实是：什么是理想时钟？对此，相对论——无论狭义相对论还是广义相对论——的回答是：理想时钟是记录自己世界线长度的时钟。这是理想时钟的定义，被托雷提称为“时钟假设”（clock hypothesis）。不难证明，对于时钟佯谬所涉及的闵科夫斯基空间的时钟来说，这一定义给出的理想时钟与瞬时随动惯性系（momentarily co-moving inertial frame）里的时钟完全同步（请读者自行证明）^[8]。由此，我们也得到了“加速度究竟会不会对时钟产生影响？”的答案，那就是加速度对理想时钟没有影响。

细心的读者也许已经注意到了，上述理想时钟的定义其实正是前面提到过的时钟佯谬现代解释的要点，即“时钟记录的是自己的世界线长度”。时钟佯谬的现代解释之所以有极大的普适性，一个很根本的原因就是它实际上包含了理想时钟的定义。

在本文的最后，给感兴趣的读者留两组思考题：

（1）人们常说的“引力场中的时钟较慢”究竟是什么意思？把它与等效原理合在一起，是否会得出与“加速度对理想时钟没有影响”相矛盾的结论？

（2）在理想时钟的定义中，只校正了加速度的影响，这是否是一种随意选择？能否把速度的影响也像加速度的影响一样校正掉？

参考文献

- [1] Misner C W, et al. Gravitation[M]. New York: W. H. Freeman, 1973.
- [2] Rindler W. Relativity: special, general, and cosmological [M]. Oxford: Oxford University Press, 2006.
- [3] Sachs R K, Wu H H. General relativity for mathematicians [M]. Berlin: Springer, 1983.
- [4] Sexl R U, et al. Relativity, groups, particles: special relativity and relativistic symmetry in field and particle physics[M]. Berlin: Springer-Verlag, 2001.
- [5] Torretti R. Relativity and geometry[M]. New York: Dover Publications, 1996.
- [6] 梁灿彬, 周彬. 微分几何入门与广义相对论 (上册) [M]. 北京: 科学出版社, 2006.

2011年5月14日写于纽约

[1] 也称为时间膨胀 (time dilation) 效应。

[2] 我试图查找丁格尔对天文学的贡献, 却没能找到。他最主要的天文活动似乎是参加了1927年与1932年的日食远征队, 但两次都因天气原因无功而返。他被选为皇家天文学会主席一事, 据说连他自己都觉得惊讶, 因为自20世纪30年代后期起, 他就已经离开天文学, 转而研究自然哲学了。

[3] 不过, 梁灿彬等人的著作多加了一个似是而非的论据, 即认为三维加速度是相对的, 四维加速度才是绝对的, 以此反驳那种认为加速度也是相对的观点。其实, 就该书所述的情形——即

该书自己援引的第6.3节——而言，在对解释时钟佯谬来说最关键的加速度的“有”和“无”的区分上，三维加速度与四维加速度都是绝对的（理由很简单，相对于一个惯性系作加速运动的物体相对于任何惯性系都是作加速运动的，从四维加速度的分量表达式也可看出，四维加速度为零当且仅当三维加速度为零），对两者作相对与绝对的划分对于解释时钟佯谬来说不仅似是而非，而且毫无必要。

[4]对于类空曲线来说，这种长度常被称为“固有长度”（proper length），对于类时曲线来说，则常被称为“原时”或“固有时”（proper time）。另外，闵科夫斯基空间常被称为“闵科夫斯基时空”。

[5]本文对“参照系”和“坐标系”这两个术语只作粗略区分：意在强调与核心物理观察者（即那两个时钟或双生子中的某一个）的关系时用“参照系”，意在强调具体数学坐标时用“坐标系”。

[6]在讨论本文的过程中，有网友提出了这样一个问题：为什么运动时钟参照系必须接受一个“不平等”的度规，而不能像静止时钟参照系那样，认为自己的度规是闵科夫斯基度规？在时钟佯谬的框架中，这是因为一开始就已假定问题发生在闵科夫斯基空间中，而所谓“静止”时钟与“运动”时钟的唯一合理的定义就是前者的世界线为测地线，后者的世界线为非测地线，而且两者都是——并且也只能是——相对于背景度规来定义的（相对论不是一个马赫式的理论，在相对论中与奥地利哲学家马赫所设想的遥远星体所起作用最接近的东西就是背景度规），这就保证了只有前者所在的参照系可以自始至终使用闵科夫斯基度规，后者则只能使用从闵科夫斯基度规（通过坐标变换）诱导出来的度规。

不过，这也引出了一个更一般的问题，那就是闵科夫斯基度规的特殊地位是从何而来的？在狭义相对论中，这可以说是一个基本假设（或经验事实）。那么，广义相对论的情况是否会强一些呢？它是否能对闵科夫斯基度规的特殊地位做出“更物理”的说明（从而也对时钟佯谬作出“更物理”的解释）呢？很遗憾，答案是否定的，因为闵科夫斯基度规的特殊地位在广义相对论中也是基本假设，因为广义相对论所用的赝黎曼空间就是局部为闵科夫斯基空间的流形（这是等效原理的体现），其度规则是可以局部地由闵科夫斯基度规诱导出来的。实际上，按照我们在正文中所建议的类比思路，闵科夫斯基度规在相对论中的地位与欧几里得度规在普通黎曼

几何中的地位是完全相似的，两者都是切空间中的度规，都是诱导其他度规的基石。广义相对论无法比狭义相对论“更物理”地解释闵科夫斯基度规的特殊地位（从而也无法“更物理”地解释时钟佯谬），就好比黎曼几何无法比欧几里得几何更充分地说明欧几里得度规的特殊地位。

[7]有人也许要问：时钟佯谬的传统解释到底算不算错误？我的看法是，在各自针对的特例或近似下，它们作为理解时钟佯谬的辅助手段，谈不上错误。但它们是否称得上解释，则取决于对“解释”一词的理解，我个人认为它们起码不算是好的解释。

[8]瞬时随动惯性系是指在所考虑的时刻与运动时钟具有相同瞬时速度的惯性参照系，也称为“瞬时静止惯性系”（momentary inertial rest frame）。

从等效原理到爱因斯坦-嘉当理论

一、等效原理

众所周知，等效原理（equivalence principle）——即引力场与加速场的不可区分性——是局域的。在一个非局域的参照系——比如有限大小的“爱因斯坦升降机”（Einstein's elevator）——中，我们可以通过对所谓“测地偏离”（geodesic deviation）效应的观测，来区分引力场与加速场。这种观测之所以有效，是因为所涉及的是联络（connection）的导数，或者说曲率（curvature）的分量，这是不能通过等效原理消去的。由于对测地偏离效应的观测是在有限大小而非局域的参照系中进行的，因此与等效原理并不矛盾。

一般教材的讨论大都到此为止。

很明显，若所有物理效应都只跟度规及联络有关，那等效原理的成立就是普遍的。但假如存在某种局域的物理效应与曲率相耦合，那么哪怕在局域的参照系中，我们也将可以通过对这种物理效应的观测，而对引力场与加速场做出区分。

那样的物理效应是否存在呢？答案极有可能是肯定的。事实上，有自旋粒子的运动很可能就是那样的物理效应之一。虽然迄今尚无任何实验足以检验这类效应，但一般认为，有自旋粒子在引力场中的运动由所谓的“马西森-帕帕佩特鲁-狄克逊方程”（Mathisson-Papapetrou-Dixon

quation) 所描述, 而这一方程显含曲率张量。因此, 有自旋粒子在引力场中的运动会与曲率相耦合。由此得出的一个推论则是, 通过观测有自旋粒子的运动, 原则上能在局域参照系中区分引力场与加速场^[1]。

从某种意义上讲, 这意味着等效原理不再成立了。

但是, 这并不意味着广义相对论失效。对于广义相对论来说, 等效原理的作用主要是确立时空的赝黎曼 (pseudo-Riemannian) 结构。为此只要在每一点上存在局域参照系, 使度规为闵科夫斯基度规 (Minkowski metric), 同时使得联络系数全部为零即可 (如果把这作为等效原理的定义, 则等效原理的成立将不受上面提到的效应所影响)。至于是否有物理效应与曲率相耦合, 并不妨碍广义相对论的建立。有自旋粒子的经典运动在广义相对论的框架中是完全可以处理的, 就像时钟佯谬在狭义相对论的框架中完全可以处理一样。

二、爱因斯坦-嘉当理论

刚才我们提到，有自旋粒子在引力场中的运动会与曲率相耦合，从而能用来局域地区分引力场与加速场。这一讨论只涵盖了与引力有关的有自旋粒子问题的一半——即有自旋粒子在给定的引力场中会如何运动。现在，我们来考虑问题的另一半，即有自旋粒子本身会产生什么样的引力场。这是一个性质很不相同的问题，因为有自旋粒子在给定的引力场中的运动——如前所述——不会对广义相对论的结构产生根本性的影响，而有自旋粒子本身产生的引力场，则——如我们即将看到的——虽非必然，却很有可能把我们引向不同于广义相对论的理论，比如爱因斯坦-嘉当（Einstein-Cartan）理论。

我们知道，对所有具有能量动量起源的角动量 $J^{abc} = x^a T^{bc} - x^b T^{ac}$ 来说，能量动量张量 T^{ab} 的守恒（即 $\partial_a T^{ab} = 0$ ）与对称（即 $T^{ab} = T^{ba}$ ）保证了角动量的守恒（即 $\partial_a J^{abc} = 0$ ）。这种角动量被称为轨道角动量，它涵盖所有的经典角动量（包括经典意义下的“自旋”——即自转角动量）。另一方面，我们也知道，并非所有的角动量都具有能量动量起源，比如量子意义下的自旋就不具有能量动量起源（因为一个有自旋粒子完全可以是无质量的）。如果我们把这种所谓“内禀”（即不具有能量动量起源）的角动量记为 S^{abc} ，则总角动量可以表示为 $J^{abc} = S^{abc} + x^a T^{bc} - x^b T^{ac}$ 。这时角动量守恒 $\partial_a J^{abc} = 0$ 将会要求

$$\partial_a S^{abc} = T^{cb} - T^{bc}$$

这一式子表明，除非内禀角动量单独守恒（即 $\partial_a S^{abc} = 0$ ），否则能量动

量张量将是非对称的（即 $T^{ab} \neq T^{ba}$ ）。由于内禀角动量显然并不单独守恒，因此上式中的能量动量张量是非对称的。

如果能量动量张量非对称，那么爱因斯坦场方程 $G^{ab}=8\pi T^{ab}$ 将要求爱因斯坦张量 G^{ab} 也是非对称的。这表明时空几何将不会是单纯的黎曼几何（Riemannian geometry）。使 G^{ab} 非对称的一种最简单的方案，就是引进非零的时空挠率（torsion） $t_{bc}^a = \Gamma_{bc}^a - \Gamma_{cb}^a$ 。由此产生的最简单的理论就是所谓的爱因斯坦-嘉当理论，是法国数学家嘉当（Élie Cartan）于1922年提出的。

与纯度规性的广义相对论不同，爱因斯坦-嘉当理论是一种建立在仿射联络（affine connection）基础上的引力理论，在这种理论中等效原理不再成立（因为非零挠率使得联络系数全部为零的局域参照系不复存在）。爱因斯坦-嘉当理论中的这种带挠率的几何被称为黎曼-嘉当几何（Riemann-Cartan geometry）。爱因斯坦-嘉当理论的场方程则为

$$G^{ab} = 8\pi T^{ab}$$

$$t_{bc}^a = 8\pi S_{bc}^a + 4\pi \delta_b^a S_{cd}^d + 4\pi \delta_c^a S_{db}^d$$

不过，上述推理并不是唯一的。

这不仅是因为使能量动量张量非对称的方法并不唯一（从而爱因斯坦-嘉当理论并不是唯一可能的推广），而且也是因为内禀角动量的出现及并不单独守恒这一特点并非必然导致能量动量张量的非对称性。事实上，通过对能量动量张量添加一个对运动方程没有影响的散度项，我们总可以将它改写为对称形式。这种对称形式的能量动量张量被称为贝林番特张量（Belinfante tensor）。有一种（比较常见的）观点认为，出现在爱因斯坦场方程中的能量动量张量应该是贝林番特张量^[2]。显然，

这可以使得爱因斯坦场方程的成立不受内禀角动量的影响。从这个意义上讲，目前并没有充分的理由——哪怕只是理论上的理由——使人们必须在经典范围内拓展广义相对论的框架。

但是，将贝林番特张量引进爱因斯坦场方程的做法也并不是完全令人满意的。比如它使得表示角动量的能量动量起源的关系式 $J^{abc} = x^a T^{bc} - x^b T^{ac}$ 具有了完全的普遍性，而我们在前面提到过，量子意义下的自旋就不具有能量动量起源。因此，角动量与能量动量之间的这种关系式似乎不该具有那么大的普遍性，起码不该将量子意义下的自旋包括在内。而一旦认定量子意义下的自旋是一种与能量动量无关的角动量，那它对时空的影响就没有理由被包含在能量动量对时空的影响——即爱因斯坦场方程——之中。

另一方面，我们也不能简单地把自旋对时空的影响从理论中丢弃掉，因为虽然尚不存在自旋对时空产生影响的任何观测证据（考虑到自旋的微小，这是不足为奇的），但由于轨道角动量对时空的影响是广义相对论的确凿推论，在理论上单单把自旋对时空的影响丢弃掉无疑是极不自然的。这些都表明爱因斯坦-嘉当理论对自旋的处理——即既承认它对时空有影响，又不把这种影响归结于能量动量——是有一定合理性的。

除此之外，爱因斯坦-嘉当理论还有其他一些值得探讨的特点，比如它可以将时空流形切空间上的结构群从广义相对论中的洛伦兹群（Lorentz group）推广到庞加莱群（Poincaré group）——这是嘉当提出这一理论的原始动机之一（我们所提及的量子意义下的自旋在当时尚未被发现），又比如它有可能对（部分地）消除广义相对论中的奇点问题起到一定帮助，等等。

不过，所有这些合理性及值得探讨的特点，都未能使爱因斯坦-嘉当理论得到太多的关注。原因在我看来有不一条：比如爱因斯坦-嘉当与广义相对论的差别涉及到了像自旋这样的量子效应，从而不仅现在，哪怕将来也几乎没有任何可能得到直接的观测支持（引力在这种尺度上太过微弱）。此外，像有自旋粒子产生的引力场那样的问题，由于场源的量子特征无法忽略，很可能根本就不能用经典理论来处理^[3]。假如经典理论根本就不能用，那么将广义相对论推广为爱因斯坦-嘉当理论的做法，也许就像当年索末菲（Arnold Sommerfeld）将玻尔理论推广为相对论性那样，缺乏真正的重要性。

2006年7月30日写于纽约

2014年12月13日最新修订

^[1]这里需要注意的是，所谓“有自旋粒子”指的是量子场论意义下的有自旋的点粒子，因为这里所借重的是量子场论意义下的“自旋”和“点粒子”这两个概念——假如所讨论的不是这种概念，而是有限大小的经典旋转物体，则与等效原理的成立与否无关（因为它不是局域的）。从某种意义上讲，这是在通过量子效应来局域地区分引力场与加速场。

^[2]支持这种观点的一个重要理由是：从引力场的作用量原理所导出的场方程自动具有对称形式的能量动量张量。对这一点感兴趣的读者可参阅拙作《希尔伯特与广义相对论场方程》的第3节——收录于本书的“姊妹篇”《小楼与大师：科学殿堂的人和事》（清华大学出版社，2014年）。

^[3]比方说，用广义相对论的克尔（Kerr）解来描述一个质量为 m ，自旋为 J 的微观粒子，将自旋视为角动量，则度规会在接近粒子康普顿波长（Compton wavelength）的 J/m 处出现所谓的“裸奇环”。我们且不去理会那个很令人头疼的“裸”字——别想歪了，这是一个技术字眼，对之感兴趣

的读者请参阅拙作《从奇点到虫洞》的第4章（清华大学出版社，2013年），在接近粒子的康普顿波长处出现像“奇环”那样的奇异性显然是不可接受的，也是与粒子物理实验完全矛盾的。虽然对微观粒子来说，我们原本就不该对经典描述有太多期待，但康普顿波长是经典与量子效应的分水岭，经典度规在“分水岭”上就出现如此巨大的问题，无疑是非常奇怪的，也是与引力在微观世界中的微弱性很不一致的。

黑洞略谈^[1]

如果要在科学术语当中评选几个最吸引大众眼球的术语，黑洞（black hole）无疑会名列前茅。这个试图用引力把自己遮盖得严严实实的家伙不仅频繁出没于科幻故事中，而且在新闻媒体上也有不低的出镜率。前不久，一条有关美国国家航空航天局（The National Aeronautics and Space Administration, NASA）的“钱德拉”X射线太空望远镜（Chandra X-ray Observatory）发现“最年轻黑洞”的新闻就被媒体竞相转载。而有关大型强子对撞机（Large Hadron Collider, LHC）有可能因产生微型黑洞而毁灭地球的传闻，更是不仅在过去几年时间里反复出现在各大媒体的显著位置上，而且还将美国和欧洲的司法界都卷入其中——因为有人试图通过法律手段来制止对撞机的启用，以“拯救”地球。在对撞机开始试运行的2008年9月，在印度甚至还发生了“一个‘黑洞’引发的血案”——一位16岁的花季女孩据说因担心微型黑洞毁灭世界而自杀。

这个搅起了如此风波的黑洞究竟是什么呢？我们就围绕这两组新闻来谈谈它吧。

黑洞这个概念的起源通常被回溯到1783年，虽然那跟我们如今所说的黑洞其实没太大关系。那一年，英国地质学家米歇尔（John Michell）利用牛顿万有引力定律和光的微粒说推出了一个有趣的结果，那就是一个密度与太阳一样的星球如果直径比太阳大几百倍，它的表面逃逸速度将会超过光速。这意味着该星球对远方观测者来说将成为一

颗“暗星”（dark star）——因为作为微粒的光将无法从它表面逃逸。不久之后（1796年），法国数学家拉普拉斯（Pierre-Simon Laplace）在其著作《世界体系》（*Exposition du système du Monde*）中也提出了同样的结果。这个如今看来只有中学水平的结果，就是黑洞概念的萌芽。

但这个萌芽很快就枯萎了。

枯萎的原因是它所依赖的前提之一——光的微粒说在科学界失了宠，被所谓光的波动说所取代。光的波动说顾名思义，就是把光看成是一种波。但牛顿引力对这种波会有什么影响？却是一个谁也答不上来的问题。既然答不上这个问题，光能否从星球表面逃逸之类的问题也就无从谈起了。因此自《世界体系》的第3版开始，拉普拉斯悄悄删除了有关“暗星”的文字，他这个“与时俱进”的做法基本上为牛顿理论中的黑洞概念画上了句号。

黑洞概念的卷土重来是在20世纪的第二个十年。那时候，爱因斯坦（Albert Einstein）于1915年底提出了广义相对论（general relativity）。1916年初，一位被第一次世界大战的战火卷到前线，且罹患天疱疮（pemphigus），“阳寿”只剩五个多月的德国物理学家施瓦西（Karl Schwarzschild）得到了广义相对论的一个后来以他名字命名的著名的解——施瓦西解（Schwarzschild solution）。从这个解中，我们可以得到很多推论，比方说如果把太阳压缩成一个半径不到3千米的球体^[2]，外部观测者就将再也无法看到阳光，这就是一种现代意义下的黑洞——施瓦西黑洞。与米歇尔和拉普拉斯的“暗星”不同，现代意义下的黑洞具有很丰富的物理内涵，并且不依赖于像光的微粒说那样的前提^[3]。

遗憾的是，施瓦西解的那些推论在很长的一段时间里不仅没有被人们所完全了解，反而遭来了一些针对黑洞的反对意见。就连爱因斯坦也

曾提出过一些如今看来很幼稚的反对意见^[4]。

不过“东边不亮西边亮”，另一个方向上的研究——即对白矮星（white dwarf）的研究——却殊途同归地将科学家们引向了黑洞。白矮星是耗尽了核聚变原料后的老年恒星，它们的质量与太阳相仿，块头却跟地球差不多，因而密度极高（一汤匙的白矮星物质的质量可达好几吨）。白矮星的发现给科学家们带来了一个问题：我们知道，恒星之所以能稳定地存在，是因为内部核聚变反应产生的巨大的辐射压强抗衡住了引力。但像白矮星那样不具有大规模核聚变反应的天体又是如何“维稳”的呢？这是一个很困难的问题。但幸运的是，当人们为这一问题伤脑筋时，一门新兴学科——量子力学——已经成熟了起来，在量子力学中有一条原理叫做泡利不相容原理（Pauli exclusion principle）。按照这条原理，电子是一群极有“个性”的家伙，每一个都坚持拥有独一无二的状态。如果你想压制这种“个性”，它们就会“殊死抗争”，这种抗争在宏观上会体现为一种巨大的压强，叫做“电子简并压”（electron degeneracy pressure）。白矮星主要就是依靠这种压强来抗衡引力的。当时很多人认为，这就是恒星的终极“养老方案”，因为计算表明，“电子简并压”在任何情况下——即对于任何质量的恒星——都足以抗衡引力。

但一位印度年轻人无情地粉碎了这个美好的“养老方案”，此人名叫钱德拉塞卡（Subrahmanyan Chandrasekhar），本文开头提到的发现“最年轻黑洞”的“钱德拉”X射线太空望远镜就是以他的名字命名的。

1930年，本科刚毕业的钱德拉塞卡在研究白矮星时发现了一个出人意料的结果，那就是如果将相对论效应考虑在内，电子简并压将大为减弱，尤其是，当白矮星的质量超过太阳质量的1.4倍时，电子简并压将无法抗衡引力。可电子简并压是当时已知的能使老年恒星抗衡引力的唯一机制，如果这一机制不管用了，那老年恒星的命运会是什么呢？这一

新问题使很多人深感不安，其中包括重量级的英国天文学家爱丁顿（Authur Eddington）。爱丁顿表示，钱德拉塞卡的结果是荒谬的，大自然是一定会让晚年恒星“老有所依”的。用今天的眼光来看，这是一种没什么说服力的单纯信念式的表态。不过在当年，这种表态却给钱德拉塞卡带来了很大的麻烦，他的论文直到一年多之后，才在遥远的美国找到一份杂志发表。

后来人们知道，恒星的“养老方案”其实不是唯一的，当电子简并压无法抗衡引力时，老年恒星还有另一种归宿，那就是中子星。这是一种密度比白矮星还高一亿倍（从而一汤匙物质的质量可达几亿吨）的天体，它依靠的是与电子简并压相类似、但更为强大的中子简并压。不过可惜的是，后者的强大也是有限度的，当中子星的质量超过太阳质量的3倍多时，中子简并压也会在巨大的引力面前败下阵来，这时的恒星就真的没救了，它的归宿只有一个，那就是黑洞^[5]。因此黑洞不仅是施瓦西解（以及后来发现的若干其他解）的推论，更是大质量恒星演化的必然归宿。

但所有这些都只是理论，接下来的问题是：像黑洞那样“黑”的东西，如何才能得到观测上的证实？答案是：“解铃还需系铃人”，能帮助我们观测黑洞的，恰恰是那个使黑洞变“黑”的幕后推手——引力。黑洞虽然不发光，它的巨大引力却足以造成许多极为显著的观测效应，比方说，如果黑洞附近有足够多的物质，甚至有大质量的伴星，黑洞的巨大引力就会吞噬那些物质，而那些物质则会在掉进黑洞之前“垂死挣扎”——因剧烈碰撞等原因而发射出强烈的X射线（图4）。探测这种X射线因此而成为了探测黑洞最重要的手段之一。

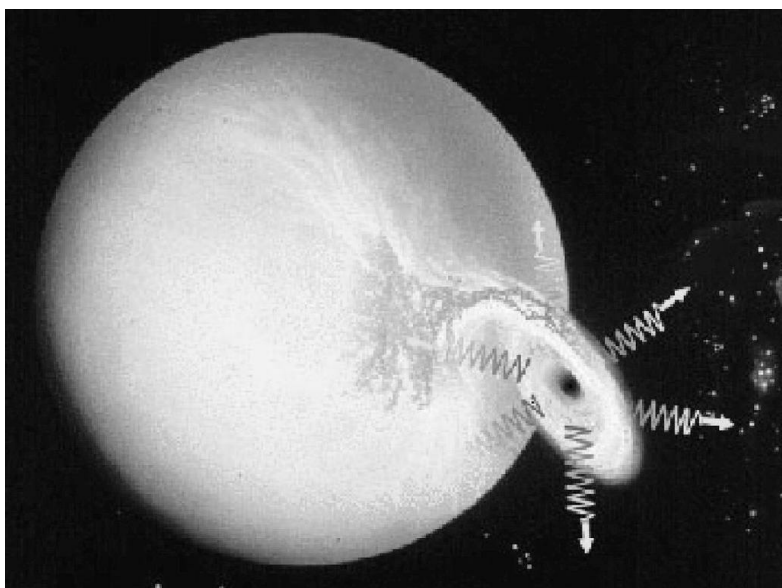


图4 黑洞因吞噬物质而发射X射线

好了，现在我们可以回过头来谈谈本文开头提到的那两组新闻了。“钱德拉”X射线太空望远镜之所以能用来寻找黑洞，正是利用了物质在掉进黑洞之前会发射出强烈的X射线这一特点。而此次发现的黑洞之所以被称为“最年轻”，是因为它只有30多岁。我们怎么知道它只有30多岁呢？因为它是1979年观测到的一次超新星爆发的遗迹。不过要补充说明的是，这个黑洞位于距我们约5000万光年之遥的一个漩涡星系中，我们如今观测到的乃是它在5000万年前所发射的X射线，因此它的真正年龄其实是约5000万岁而不是30多岁。我们又怎么知道它是黑洞呢？那是因为天文学家们利用X射线能谱等数据估算了它的质量，结果约为太阳质量的5~10倍，超过了中子星的最大可能质量^[6]。这就是“最年轻黑洞”这一头衔的由来。

接下来再谈谈所谓大型强子对撞机有可能因产生微型黑洞而毁灭地球的传闻。大型强子对撞机是一个设计能量为7万亿电子伏特（7TeV）的对撞机（图5）。那样的对撞机会产生黑洞吗？按照广义相对论，答案是否定的。因为这种万亿电子伏特（TeV）量级的能量在微观上虽然

很高，用宏观标准来衡量却是微乎其微的，只不过是千万分之一焦耳的量级，这一丁点儿能量若想形成黑洞，除非是把它压缩到一个线度为一千亿亿亿亿亿分之一米（ 10^{-51} 米）的区域内，这比所谓的普朗克长度（Planck length）还小得多，与大型强子对撞机所能触及的最小线度相比，更是只有后者的一亿亿亿亿分之一（ 10^{-32} ）。因此按照广义相对论，大型强子对撞机是绝不可能产生微型黑洞的。



图5 大型强子对撞机

既然如此，为什么仍有那么多人担心微型黑洞呢？因为他们背后有“军师”在指点，那些“军师”为他们的担心注入了一条重要理由，那就是在某些现代物理理论——比如超弦理论（superstring theory）——中，时空有不只四个维度。由于引力与时空密切相关，因此时空若有不止四个维度，引力的规律也将有所不同，而引力的规律一旦不同，产生黑洞的条件就会发生变化。理论计算表明，在那些带有额外维度的理论

中，确实存在一些尚未被实验所排除的参数范围，使得大型强子对撞机有可能产生黑洞。

这么一来，事情就不太妙了。虽然那些认为时空有不止四个维度的理论目前还都只是假设性的，而那些使大型强子对撞机产生黑洞的参数范围更是假设中的假设。但无可否认的是，有不少物理学家对那样的理论寄予厚望。因此，那样的理论所允许发生的事情即便只是假设性的，也不容忽视。毕竟，我们只有一个地球，实在不敢拿她去赌哪怕最细微的风险。

幸运的是，即便那些假设性的理论是正确的，并且参数也恰好处在能使大型强子对撞机产生黑洞的范围内，那样的黑洞依然是不可能毁灭地球的。因为黑洞还有一个我们尚未介绍的重要特点，那就是它并不是完全“黑”的。1974年，英国物理学家霍金（Stephen Hawking）发现，由于量子效应的影响，黑洞会向外辐射能量。这种所谓的霍金辐射（Hawking radiation）对于大质量黑洞来说是微乎其微的，但对微型黑洞却极为显著，而且在时空有不止四个维度的情况下依然存在。计算表明，由于霍金辐射的存在，即便大型强子对撞机能够产生黑洞，那些黑洞也会在瞬息之间就“人间蒸发”，别说毁灭地球，就连侵吞一两个原子都未必来得及。

至此，大型强子对撞机有可能因产生微型黑洞而毁灭地球的传闻似乎该烟消云散了。但事实却不然，有些人依然表示怀疑，因为霍金辐射尚未被观测证实过。虽然有关微型黑洞毁灭地球的担心本身也是建立在尚未被观测证实过的理论之上的，但当科学家们用同样类型的理论来回答他们的担心时，有些人却拒绝接受。对于这种近乎偏执的怀疑，有一样东西可以替科学家们作出回应，那就是宇宙射线。

大型强子对撞机是人类迄今所建能量最高的对撞机，但浩瀚的宇宙却有各种办法产生比那高得多的能量。观测表明，我们所栖居的地球每秒钟都会受到10万次以上的超高能宇宙射线的轰击，那些宇宙射线与地球物质发生碰撞时所具有的能量比大型强子对撞机的能量更高^[7]，而且那样的轰击自地球诞生以来，在长达45亿年的时间里从未间断过，相当于每时每刻都有大型强子对撞机在运行。如果大型强子对撞机果真有产生微型黑洞并毁灭地球的风险，无论其理论机制是什么，那样的风险都早该被宇宙射线转化为现实了。我们今天仍能坐在地球上争论这一问题本身，就很好地说明了那样的风险并不存在。事实上，如果我们把眼光放得更远一点，那么不仅地球每时每刻都受到大量超高能宇宙射线的轰击，表面积是地球一万多倍的太阳更是一个大得多的靶子，如果那样的轰击有危险的话，像太阳那样的庞然大物无疑会比地球死得更快。因此，包括太阳在内所有恒星的存在全都是极强的证据，表明大型强子对撞机因产生微型黑洞而毁灭地球的风险是完全可以排除的^[8]。

事实上，大型强子对撞机若果真能产生微型黑洞的话，那不但是不是什么风险，反而是了不起的实验成就，因为那不仅是对某些现代物理理论的绝佳检验，而且还是研究霍金辐射的最好、甚至有可能是唯一的直接手段。

2010年11月24日写于纽约

[1]本文曾以《有关黑洞的前世今生》为题发表于《中学生天地》2011年2月刊（浙江教育报刊社出版）。

[2]更准确的说法是周长不到18.6千米（3千米 $\times 2\pi$ ），因为那才是具有观测意义的量。但为行文方便起见，我们仍将使用“半径”这一术语，只不过它的真正含义是周长除以 2π ，而非径向距

离。

[3]现代意义下的黑洞（施瓦西黑洞只是其中最简单的一种）与米歇尔和拉普拉斯的“暗星”很不相同，比如后者只是远方的观测者无法看到（由于作为微粒的光在“暗星”引力场中仍可运动一段距离，因此近处的观测者仍可看到），而前者则对于任何外部观测者都是“黑”的。

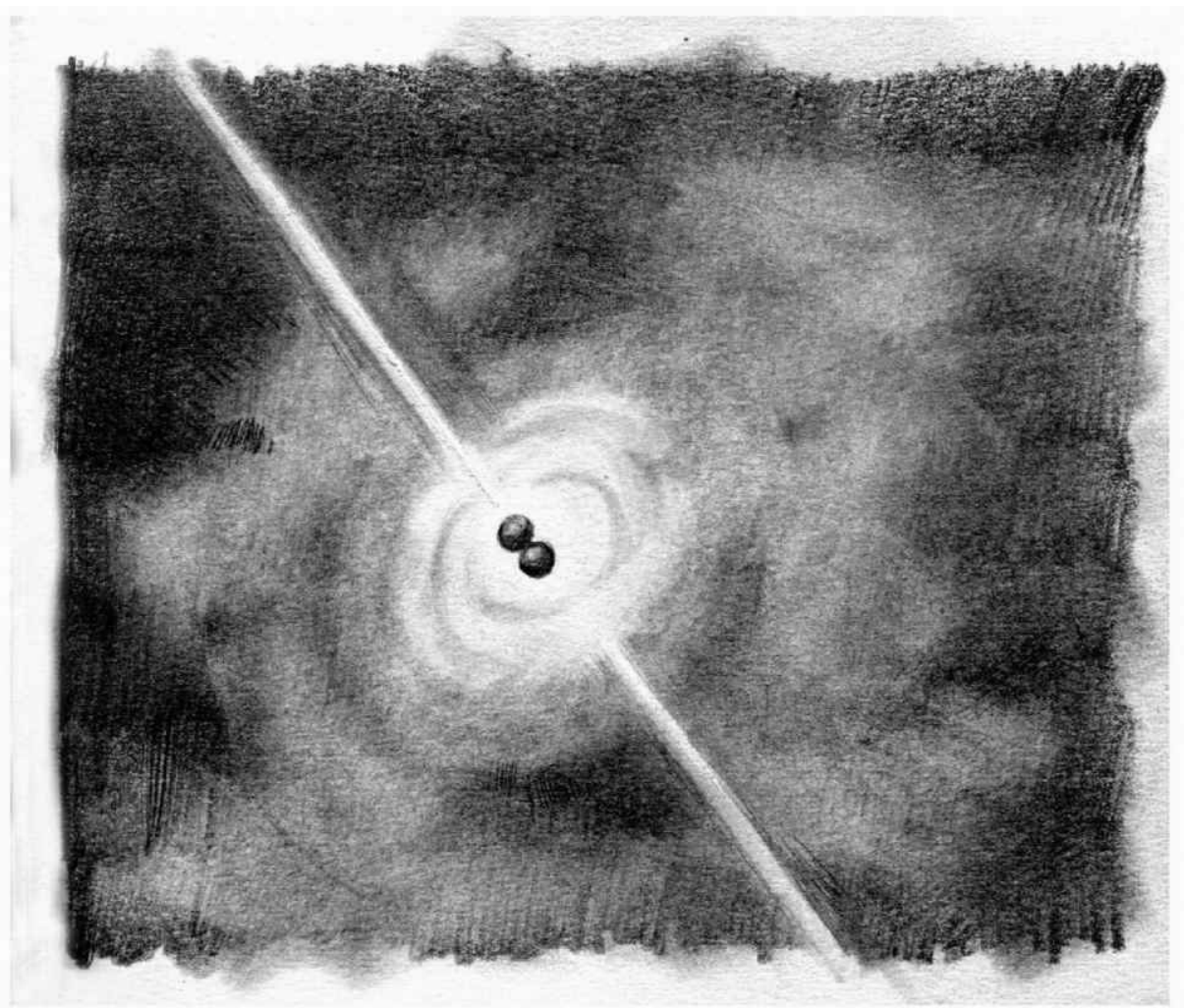
[4]爱因斯坦计算了黑洞附近圆轨道上的粒子运动速度，结果发现轨道半径小于黑洞临界半径的1.5倍时，粒子运动速度会超过光速。他据此认为黑洞是不可能存在的。这一意见的幼稚之处在于，那计算无非说明在黑洞近旁粒子不可能维持圆轨道（除非有外力），而并不表示黑洞无法存在。这就好比在一个大漩涡里游泳者无法维持圆轨道，并不表示大漩涡不可能存在。

[5]有人提出过比中子星更致密的所谓“夸克星”（quark star）。不过“夸克星”即便存在，其密度也只会比中子星略大（如果说中子星像一个巨型原子核，那么夸克星就像一个巨型核子）。“夸克星”是否存在目前尚有争议，不过理论研究显示，无论它存在与否，都不太可能显著改变耗尽核聚变能量后大质量天体坍缩为黑洞的临界值。

[6]不过，由于对中子星最大可能质量的计算以及对“最年轻黑洞”的质量估算都有一定的误差，因此该天体究竟是黑洞还是中子星目前尚有一定的争议，只能说它有较大的可能性是黑洞。

[7]这个能量是指质心系能量。

[8]严格讲，由高能宇宙射线产生的微型黑洞——如果有的话——与大型强子对撞机产生的微型黑洞有一个区别，那就是前者是高速运动的，从而会很快穿过地球。但研究表明，即便如此，假如那样的微型黑洞能够被产生，并且有毁灭星球的威力的话，宇宙中那些高度致密且具有强引力场的天体——比如白矮星和中子星——仍会因为俘获那样的黑洞而迅速灭亡，这同样与观测明显不符。



绘画：张京

反物质浅谈

一、一个令人苦恼的结果

众所周知，科幻小说作为一种特殊形式的小说，常从现代科学的发展中吸取新概念，反物质就是常被吸收的新概念之一。20世纪40年代，美国科幻小说家威廉森（Jack Williamson）创作了一系列以反物质为题材的小说，称为C.T.故事，其中“C.T.”是他为反物质所拟的名称——“Contra-Terrene”——的缩写。威廉森的C.T.故事问世后不久，另一位美国科幻小说家阿西莫夫（Isaac Asimov）也在自己脍炙人口的机器人故事中引进了反物质的概念，他所设想的机器人脑是所谓的“正电子脑”（positronic brain），而正电子乃是电子的反粒子，是反物质的基本组元之一。20世纪60年代，著名科幻电视连续剧《星际迷航》（*Star Trek*）开始播出，在这部连续创作和播出约40年之久、拥有不止一代忠实粉丝的电视连续剧中，反物质是星际飞船的重要燃料。这一点如今已几乎成为了所有以星际旅行为题材的科幻小说的共同特点。反物质概念在科幻小说中的频频出现，使公众对这一概念也产生了浓厚兴趣。

那么，反物质这一概念是何时，以何种方式被提出的？人们又是如何发现反物质的？反物质究竟是不是一种有效的星际飞船燃料？我们的宇宙中到底是物质多呢还是反物质多？这些或许是很多人不甚了解却不无兴趣的问题。本文将对这些问题作一些介绍。

反物质这一概念在学术界的出现最早可以追溯到19世纪末。1898年，英国物理学家舒斯特（Arthur Schuster）在给《自然》（*Nature*）杂志的一封信中提到，既然电荷可以有负的，金子说不定也可以有负的，而且负金子说不定和我们熟悉的金子有着一样的颜色。这或许是有关反物质的想法在科学文献中的萌芽。不过舒斯特有关反物质的想法只是一种简单而模糊的思辨，没有真正的理论依据，因而也没有引起任何重视。反物质概念在物理学上的真正渊源，是从将近30年后的1927年开始的。那一年，量子力学奠基人之一的英国物理学家狄拉克（Paul Dirac）提出了一个描述电子运动的数学方程。

狄拉克所提出的这一方程——即所谓的狄拉克方程（Dirac equation）——是一个既具有量子力学特征，又满足狭义相对论要求的方程，在当时是很令人耳目一新的结果^[1]。更漂亮的是，这一方程还出人意料地自动包含了一些此前为解释实验结果而不得不人为添加到量子力学中的东西，一些在当时看来绝非显而易见的东西，比如电子的自旋和磁矩。作为一个方程式，狄拉克方程的形式之简洁，内涵之丰富，预言之神奇，似乎达到了物理学家们梦寐以求的境界。

但这一方程的“野心”似乎还不止于此，它还包含了另外一个重要结果——可惜这回却是一个令人苦恼的结果。

这个令人苦恼的结果是：狄拉克方程所描述的电子的总能量既可以是正的，也可以是负的。这个结果之所以令人苦恼，是因为人们在自然界中从未发现过总能量为负的电子，因此狄拉克方程似乎允许存在一些自然界中不存在的东西。仅仅这样倒还罢了，因为允许存在的东西可以碰巧不存在，因此大不了假定自然界中所有电子的总能量碰巧都是正的。但不幸的是，按照量子力学，一个理论只要允许总能量为负的状态——即所谓的“负能量状态”，那么哪怕假定自然界中所有的电子的总能

量碰巧都是正的，它们也会在很短的时间内通过量子跃迁进入到负能量状态，从而变成总能量为负的电子——也称为“负能量电子”。这种跃迁的结果无疑是灾难性的，与现实世界也大相径庭^[2]。

二、错误描述中的正确结论

这么看来，狄拉克方程看似漂亮，实际上却似乎是错的，而且还错得相当离谱，足可把整个世界都搭进灾难里去。但是，狄拉克方程又分明包含了很多看起来正确得惊人的结果，一个错得如此离谱的方程又怎么可能包含如此多正确得惊人的结果呢？莫非真的应了那句俗语：真理过头一步就是谬误？

为了解决这个令人苦恼的两难问题，狄拉克于1930年提出了一个大胆的假设，那就是负能量电子的确是存在的，不仅存在，而且还很多，多到足以把所有负能量状态都占满的地步。有人也许会问：既然有这么多个负能量电子，为什么人们在自然界中从未发现过呢？答案是：由所有这些负能量电子组成的“海”就是我们平时所说的真空，从而不存在直接的观测效应。狄拉克之所以提出这样古怪的假设，是因为当时人们已经知道了一条重要的物理原理，叫做泡利不相容原理（Pauli exclusion principle），它表明任何两个电子都不能有相同的状态。既然任何两个电子都不能有相同的状态，那么一旦所有负能量状态都被负能量电子所占满，正能量电子也就不可能再通过量子跃迁进入到负能量状态了。这样一来，负能量状态的存在也就不再成为问题了。

狄拉克的假设挽救了狄拉克方程，却带来了一个新问题。那就是他的假设虽然阻止了正能量电子进入负能量状态，却并不妨碍负能量电子因获得外来的能量而变成正能量电子。一旦出现这种情形，除产生一个正能量电子外，真空中还将出现一个因负能量电子空缺而形成的空穴，这种空穴等价于一个具有正能量，并且带正电荷的粒子。（请读者想一想这是为什么？）由此带来的新问题就是：这种带正电的粒子究竟是什么？

么粒子呢？狄拉克的数学直觉告诉他那应该是一个质量与电子质量相同的粒子。但当时物理学家们所知道的唯一带正电的基本粒子是质子，其质量比电子质量大了1 800多倍。因此如果空穴所对应的带正电粒子的质量与电子质量相同，它将是一种新粒子，这是一个很大的麻烦。今天的读者也许难以理解这种视新粒子为麻烦的想法，因为换作是在今天，能够预言新粒子不仅不是麻烦，往往还会被认为是令人兴奋的结果（除非有显著的实验证据或理论依据表明所预言的新粒子不可能存在）。但提出新粒子这种后来一度成为家常便饭甚至蔚为时尚的做法，对当时的物理学家来说却几乎是一个思维禁区——一个连素以勇气著称的量子力学奠基者们也未敢轻易逾越的思维禁区。在这一思维禁区面前，具有极高数学天赋，并且一向崇尚数学美的狄拉克犯下了一生为数不多的显著错误之一，他放弃了自己的数学直觉，提出空穴对应的粒子是质子。

幸运的是，思维禁区束缚得了思维，却束缚不了计算；物理学家的思维禁区束缚得了物理学家，却束缚不了数学家。狄拉克的观点提出后，与他同时代的德国物理学家海森伯（Werner Heisenberg）和奥地利物理学家泡利（Wolfgang Pauli）分别对空穴的质量进行了计算，结果表明它应该与电子质量相同；德国数学家外尔（Hermann Weyl）更是从理论的对称性出发直接证明了这一点。另一方面，不管空穴是什么，既然它是电子离开所留下的，那么电子显然也可以重新跃回空穴，一旦出现这种情况，电子与空穴就会一起消失（变成能量），这种过程被称为湮灭（annihilation）。如果空穴是质子，那么这就意味着电子可以与质子互相湮灭。这结果看起来显然很令人不安，因为电子和质子是组成物质的基本粒子（当时中子尚未被发现），如果它们可以相互湮灭，那么物质的稳定性就成问题了。当然，问题到底有多严重还得看湮灭的快慢程度，或者说湮灭的几率。美国物理学家奥本海默（Robert Oppenheimer）和俄国物理学家塔姆（Igor Tamm）分别计算了这种几

率，结果发现它相当大，足以使物质世界在很短的时间内就崩溃离析。

在这些结果的连环打击下，空穴是质子的假设遭到了灭顶之灾。1931年，狄拉克纠正了自己的错误，并提议将空穴所对应的质量与电子质量相同，电荷与电子电荷相反的实验上尚未发现的新粒子称为反电子（**anti-electron**）。这一回，他彻底突破了禁区，不仅提出了反电子，而且进一步提出质子及其他粒子——如果有的话——也应该有相应的反粒子。

如果所有的粒子都有反粒子，那么就完全有可能存在由反粒子组成的物质，这种物质就是人们所说的反物质。因此从某种意义上讲，这一年——即1931年——可以被视为是反物质概念诞生的年代。

按照狄拉克对反粒子的描述，反粒子是粒子脱离负能量状态后留下的空穴，因此反粒子与相应的粒子可以湮灭。这种湮灭有可能使粒子与反粒子同时转化为能量（比如光子）^[3]，这是理论上所能达到的最高能量转化效率。这种转化效率是如此之高，以至于1克反物质与1克物质湮灭所产生的能量就足以超过“二战”末期美军投掷在日本广岛和长崎的两颗原子弹所释放能量的总和。不难设想，若有朝一日人类能广泛利用反物质作为能量来源，无疑将会带来巨大的技术飞跃。这是反物质成为很受科幻小说家们青睐的能量来源的根本原因。

不过需要指出的是，狄拉克对反粒子的描述虽然很直观，并且粗看起来颇有道理，在今天看来其实却只有历史价值，或者用美国物理学家许温格（**Julian Schwinger**）的话说，是“最好作为历史的猎奇而被遗忘”。为什么呢？因为如上文所介绍，狄拉克的描述需要通过泡利不相容原理来阻止正能量粒子进入负能量状态。对于电子和质子这样的粒子——被称为费米子（**fermion**）——来说，这恰好是可以做到的。但自然

界中还存在另外一类粒子——被称为玻色子（**boson**），它们并不满足泡利不相容原理。对于那样的粒子，狄拉克有关反粒子的描述就无能为力了。不仅如此，按照狄拉克的描述，正反粒子的产生必须是成对的，因为一个新粒子的产生必定会留下相应的空穴——即它的反粒子；反过来说，新空穴的出现也只能是由于相应粒子的产生——即脱离负能量状态。但实验却表明这种粒子与相应反粒子的“双宿双飞”并不普遍成立。比方说在 β 衰变中，电子的出现就并不伴随有反电子。因此狄拉克对反粒子的描述细究起来并不正确，这一点不仅被多数科普读物所忽视，甚至在一些现代教科书中都没有明确说明，这是很有些不应该的。对反粒子的普遍描述，是在量子场论出现之后才建立起来的。不过狄拉克对反粒子的描述虽然并不正确，其所包含的一些基本结论，比如反粒子与相应的粒子质量相同，所带电荷及若干其他量子数相反，正反粒子可以相互湮灭，等等，却是普遍成立的，并且它的提出对量子场论的产生起到过启发作用，从这些意义上讲它对物理学的发展是功不可没的。

三、走错方向的电子还是走对方向的正电子？

与反粒子理论的曲折发展同样生动坎坷的，是实验物理学家们发现反粒子的故事。对于实验物理学家们来说，这个故事多少带着点遗憾，因为其实早在狄拉克提出反粒子概念之前，反粒子就已经在实验室里留下了踪迹，却被他们所忽略，这才让理论物理学家捷足先登。

在20世纪30年代，物理学家们探测带电粒子径迹的主要工具是云室（cloud chamber）。云室不仅可以显示带电粒子的径迹，通过将其置于磁场中，还可以进一步判断出粒子所带电荷的正负——因为正电荷与负电荷在穿过磁场时会往不同方向偏转。早在狄拉克提出反粒子概念之前，实验物理学家们就在云室照片中发现过一些类似于电子，却与电子有着相反偏转方向的径迹。这些径迹其实正是反电子掠过云室留下的倩影。可惜就像狄拉克起初不敢把空穴诠释成反电子一样，实验物理学家们也未曾想到把那些反常径迹诠释成新粒子，从而错失了先于理论而发现反电子的机会。

直到狄拉克提出空穴是反电子之后，云室中那些反常径迹才引起了一些实验物理学家的重视。比如英国卡文迪许实验室（Cavendish Laboratory）的物理学家布莱克特（Patrick Blackett）就告诉狄拉克说，自己与同事可能已经发现了反电子存在的证据。但即便有狄拉克当出头鸟，布莱克特仍未敢贸然发表自己的发现，而是打算做进一步的核实。这一延缓将发现反电子的优先权拱手让给了大西洋彼岸的美国物理学家安德逊（Carl David Anderson）。

安德逊当时在美国西岸的加州理工大学（California Institute of Technology）从事宇宙射线研究。与其他一些实验物理学家一样，他也

在自己的云室照片中发现了类似于电子，却与电子有着相反偏转方向的径迹，而且这样的径迹并不稀少，这一点引起了安德逊的重视，于是他把这一发现告诉了当时正在欧洲进行访问的导师密立根（Robert Andrews Millikan）。密立根是一位实验物理大师，曾因测量电子电荷及光电效应方面的工作获得1923年的诺贝尔物理学奖。对于安德逊所发现的径迹，密立根的解释是视之为质子产生的——质子所带电荷与电子相反，因而可以解释观测到的偏转方向与电子相反这一事实。但密立根的质子解释有一个致命的弱点，那就是像质子这样的重粒子在云室中的径迹应该远比像电子那样的轻粒子来得显著。可是安德逊所发现的径迹却并未显示出这种差异，因此密立根的质子解释很快被排除了。

另一方面，安德逊自己也提出了一种解释，他认为偏转方向与电子相反的径迹有可能是由反方向运动的电子产生的，这种解释也曾被欧洲物理学家们采用过。单纯从径迹的偏转方向上讲，它的确是能够说得通的。但安德逊的反向电子解释也有一个令人困惑的地方，那就是他所研究的是宇宙射线，而宇宙射线来自天空，从而应该是以大体相同的方向——即自上而下——穿越云室的。既然如此，反方向运动的电子又从何而来呢？解决这一疑问最直接的办法无疑是对电子的运动方向进行直接检验。为此，安德逊在自己的云室中间插入了一片薄薄的铅板。由于粒子穿过铅板速度会变慢，因此只要对粒子在铅板上下的速度快慢进行比较，就可以判断出粒子的运动方向^[4]。通过这一手段，安德逊发现绝大多数偏转方向与电子相反的粒子和电子一样来自天空，也就是说它们的运动方向与电子是相同而不是相反的。这就把安德逊自己的反向电子解释也排除了。

这两种解释都被排除了，留给安德逊的就只剩一种解释了，那就是：他所发现的径迹来自一种带正电的、质量却远比质子轻的粒子——

一种尚不被实验物理学家所知道的新粒子。但这种解释也有一个问题：那就是这样一个质量不大的新粒子为什么以前一直未被发现呢？如果安德逊知道狄拉克的空穴理论，他或许会想到那是因为这种粒子是反电子，它很容易因为与电子相互湮灭而从人们眼皮底下消失。可当时安德逊并不知道狄拉克的空穴理论，因此留给他的这唯一解释似乎看起来也不太可能。不过“看起来不太可能”和“不可能”终究是有差别的，福尔摩斯有一句虽不严谨但很管用的名言：当你排除了所有的不可能，剩下的无论看起来多么不可能，一定就是真相。安德逊知道这时候不应该犹豫了，于是他不顾密立根的反反对，于1932年9月公布了自己的发现。

4年后，这一发现为他赢得了诺贝尔物理学奖。

安德逊发现新粒子的消息一传到欧洲，布莱克特和他的同事立刻意识到自己犯下了迟疑不决的“兵家大忌”，他们已经发现却未敢贸然发表的显然正是同样的粒子。于是他们立刻也发表了自己的结果。他们的结果虽不幸在时间上落后于安德逊，却有幸在空间上占据了一个有利条件，那就是他们离狄拉克很近。安德逊虽然发现了新粒子，却不知道它和电子的关系，而布莱克特和他的同事不仅知道新粒子和电子的关系，还知道它和电子可以成对产生，于是他们在自己的云室照片中有意识地寻找这种产生过程的证据，并如愿以偿地成为了首先发现正反粒子对产生过程的物理学家^[5]。

在这些成果的发表过程中，反电子获得了一个新的、后来更为流行的名称：正电子（positron）。这个名称是一位杂志编辑向安德逊建议的，它的本意是“正子”（当时安德逊并不知道这一粒子与电子有关）。

四、从反粒子到反物质

正电子成为人类发现的第一种反粒子并非偶然。因为与之相比，其他反粒子要么在宇宙线及天然放射源中比较稀少，而早期加速器的能量又不足以产生；要么由于相互作用太弱而不易检测，其发现的难度都远远大于正电子。因此自正电子被发现之后，发现反粒子的步伐停顿了下来，直到二十几年后才迎来了一轮爆发。1955年，意大利物理学家赛格雷（Emilio G. Segrè）与美国物理学家张伯伦（Owen Chamberlain）“领衔”发现了反质子（赛格雷和张伯伦获得了1959年的诺贝尔物理学奖）；次年，美国物理学家考克（Bruce Cork）及其合作者又发现了反中子。至此，组成物质的三种最重要粒子的反粒子都被发现了。此后，随着加速器能量的持续提高，其他基本粒子的反粒子也被陆续发现——当然，后来的那些发现对物理学家们来说已毫无悬念，因为在理论上，除少数粒子与自己的反粒子相同外，所有其他粒子都该有自己反粒子的观念早已确立。

不过尽管反粒子的发现和产生已不再稀罕，但反粒子很容易被“正”粒子湮灭，因此如何保存它们依然是一个极大的技术难题。直到20世纪80年代，物理学家们才开始掌握了保存少量反粒子的手段。但是要想保存更多的反粒子，却又面临另一个技术难题，因为带同种电荷的反粒子相互排斥，中性的反粒子又不稳定。在这种情况下，要想积累反粒子，一种可能的手段是让反粒子像普通粒子配成原子那样配成中性的反原子。但是让那些极易湮灭，通常又高速运动的反粒子乖乖地组成原子又谈何容易？这项工作直到1995年才由德国物理学家欧勒特（Walter Oelert）领导的实验小组所完成，他们在欧洲核子中心（CERN）的低能反质子环（Low Energy Antiproton Ring）上成功地制备出了9个反氢原

子。虽然只有区区9个，与普通原子动辄就是几个摩尔——1摩尔约有6000万亿亿（ 6×10^{23} ）个——的海量相比少得简直不值一提，但这一消息1996年初一经披露立即引起了世界性的轰动。许多大媒体用显著标题进行了报道，欧勒特本人也受到了媒体记者的“围追堵截”，有记者甚至试图把他从飞机上拦截下来进行采访。反氢原子的制备之所以引起媒体如此广泛的关注，一个很重要的原因是因为原子和分子是承载物质物理和化学性质的基本组元。从这个意义上讲，反氢原子的成功制备是人类有史以来首次制备出了反物质，此前所研究的只能称为是反粒子而不是反物质。对媒体来说，这无疑是一个极大的兴奋点。

不过欧勒特制备反氢原子虽是欧洲核子中心有史以来最受媒体关注的新闻之一，但该中心的粒子物理学家们却大都只是将之视为实验工艺上的成就，有人甚至戏称其为“新闻实验”。因为从理论上讲，由反粒子组成反原子乃是稀松平常之事；而从实用的角度讲，欧勒特制备的反氢原子不仅数量稀少，而且存在的时间也短得可怜，只有一亿分之四秒

（ $4 \times 10^{-8} \text{s}$ ），距离实用无疑还差得很远。欧勒特实验成功后的第二年，欧洲核子中心关闭了为这一实验及其他三十几个实验立下过汗马功劳的低能反质子环。这个低能反质子环在它服役的14年间总共产生了超过100万亿个反质子。如果把这些反质子全部当成反物质燃料与质子湮灭，它们所产生的能量大约可以让一盏100瓦的灯泡点亮5分钟。将这点微不足道的能量与14年间为产生这些反质子而消耗的巨大能源相比，不难看到用反物质作为能源在目前还是极度得不偿失的。

但这些技术上的困难并不妨碍人类的想象力将反物质作为未来可能采用的一种能源。这种能源除了具有理论上最高的转化效率外，还有一个非常吸引人的优势，那就是洁净。我们知道，传统的能源，无论是化学能还是核能，通常都会在使用后产生有害的残留物，比如废气、核废

料等，而正反物质的湮灭却可以将燃料彻底转化为能量，从而不留下任何残留物质，因此它是一种理论上最洁净的能源。这样既洁净又高效的能源不仅是科幻小说家的最爱，对于工程和军事领域来说也有着无穷的魅力。比如早在20世纪中叶，美国氢弹之父泰勒（Edward Teller）和苏联氢弹之父萨哈洛夫（Andrei Sakharov）就各自提出过反物质武器的可能性。在美苏冷战的后期，伴随“星球大战”计划的展开，美国军方开始了反物质应用方面的研究。

不过，反物质武器的制造除了有上面提到的困难外，还会面临一个意想不到的难题，那就是正反物质相互接触时，因湮灭而产生的辐射压会将正反物质剧烈推开，从而急剧减缓能量释放的速度。这种效应的一个“日常生活版”很多人也许早已见过，那就是：将一滴水滴在热锅上，水会渐渐蒸发，一般来说，锅越热，蒸发就越快，可是当锅热到一定程度后，水滴的蒸发状况会发生显著变化，它会在热锅上四处移动甚至跳跃，蒸发速度则反而大为减缓。这种有趣的现象早在两百五十多年前就被一位名叫雷登弗罗斯特（Johann Gottlob Leidenfrost）的德国医生注意到了，因而被称为雷登弗罗斯特效应（Leidenfrost effect）。雷登弗罗斯特效应的物理机制是：当锅热到一定程度后，水滴剧烈汽化产生的蒸汽会在水滴与锅之间产生一层蒸汽膜，阻隔两者的进一步接触，从而急剧减缓水滴的蒸发速度。这种机制也适用于正反物质的接触，只是蒸汽膜换成辐射层而已。雷登弗罗斯特效应对反物质武器的制造是一种障碍。不过，随着苏联的解体和冷战的落幕，近乎军事“大跃进”的反物质武器研究本就很快遭到了放弃。

到目前为止，除了基础物理研究外，反物质的主要应用领域是在医学影像方面。由于技术水平及反物质数量的稀少所限，多数其他类型的反物质应用起码在目前还是很现实的。不过，让想象力自由驰骋的

话，未来的希望总是有的。比方说，假如宇宙中存在足够规模的天然反物质源，情况就将有所不同，因为那样我们就不必为制备反物质而费心了——虽然高效而安全地收集和保存反物质仍将是极具难度的挑战。

这就给科学家们提出了一个很大的问题，那就是：宇宙中有可能存在大规模的天然反物质源吗？

五、宇宙的主人和客人

物理学家们曾经对这一问题作出过肯定的猜测。狄拉克在他的诺贝尔演讲中就曾表示，如果正反物质是完全对称的，那么宇宙中完全有可能存在由反物质组成的星球。如果将这种猜测发挥一下，那么我们还可以设想宇宙中不仅存在由反物质组成的星球，甚至有可能存在由反物质组成的生物。另一方面，在宇宙大爆炸初期的极高温条件下，正反物质的产生应该是同等可能的，从这个角度讲似乎也有理由预期宇宙中存在大量的反物质，甚至在数量上与物质等量齐观。

但随着理论和观测的逐步深入，这些初看起来不无合理性的猜测渐渐冷了下来。

首先可以明确的一点是：由于反物质与物质会相互湮灭，因此在我们所生活的这颗小小的蓝色星球上，像发现煤矿或铀矿那样发现“反物质矿”是完全不可能的。不仅如此，反物质在整个太阳系中的存在也是微乎其微的，因为否则的话，由太阳发出，被称为太阳风的粒子流与反物质之间的湮灭早就应该被发现了。再往远处看，情况也没有实质的改变，虽然宇宙射线中存在一定数量的反粒子，有些地方甚至存在反粒子源，但那些反粒子大都来自普通物质所参与的高能物理过程。迄今为止并无任何确凿的证据，表明宇宙中可能存在反物质星球，或任何其他大范围的反物质分布。

事实上，不仅没有确凿证据表明宇宙中存在大范围的反物质分布，相反，却有不少证据表明大范围的反物质分布不太可能存在。这种证据之一来自于宇宙中重子——主要是质子和中子——数量和光子数量的比值。我们知道，极早期宇宙中充斥着各种基本粒子，它们随时被高能物

理过程所产生，也随时相互湮灭。当宇宙的温度逐渐降低时，粒子的产生过程开始受到抑制，因为它们所需的能量越来越难以达到。对于重子和反重子来说，这大致发生在宇宙温度为10万亿度的时候。在这个温度以下，湮灭过程起到主导作用，重子与反重子很快因为彼此湮灭而转变为光子或其他轻粒子。在那样的过程中重子与反重子变得越来越少，直至其密度低到连湮灭过程也无法有效进行为止，那时仍残留的重子就组成了我们今天所生活的物质世界（由此可见我们的物质世界是多么地来之不易）。这种过程所导致的一个显而易见的后果，就是今天宇宙中的重子数远远少于光子数，而且早期宇宙中的重子与反重子越对称，这种湮灭过程就会进行得越彻底，今天宇宙中的重子数相对于光子数也就会越少。观测表明，今天宇宙中的重子数与光子数的比值大约为1比10亿（ 10^{-9} ）。这虽然已经是一个很小的比例，但理论计算表明，如果湮灭过程开始起主导作用时宇宙中的重子与反重子是完全对称的话，这个比值还要小得多，大约会是1比100亿亿（ 10^{-18} ）。因此，我们所观测到的重子数与光子数的比值是一个很有力的证据，它表明早期宇宙中的重子与反重子是不对称的，而我们赖以生存的整个物质世界正是这种不对称的产物，是一个反物质极为稀少的宇宙。

有读者可能会问，是否有可能出现这样的情况，即早期宇宙中的重子与反重子完全对称，只不过由于某种原因而彼此分离开来，从而没有发生有效的相互湮灭？如果是这样，那就既可以保持物质与反物质之间的对称性，又可以解释为什么我们观测到的重子数与光子数的比值远比由对称性所预期的1比100亿亿来得高。应该说，这是一个很不错的问題，事实上，物理学家们曾经考虑过这样的可能性。但这种猜测有两个致命的弱点：一是没有任何已知的物理过程可以将随机产生的重子和反重子有效地加以分离；二是如果早期宇宙中真的存在过这种正反物质分离的情况，那么正反物质的湮灭在空间分布上将是高度非均匀的，这应

该会在今天的宇宙微波背景辐射中留下遗迹。这样的遗迹并未被发现，因此这种可能性基本上可以被排除了。因此，无论观测还是理论都表明：我们今天所生活的宇宙是一个正反物质不对称的宇宙，物质是这个宇宙的主人，反物质只是稀客。

六、恼人的不对称之谜

既然我们所生活的宇宙是一个正反物质不对称的宇宙，那么一个很自然的问题就产生了，那就是为什么会出现这种不对称？对此，科学家们曾经有过两类不同的看法。其中第一类看法认为正反物质的不对称是由初始条件决定的，或者说是“先天”造就的。显然，这类看法比较消极，几乎等于是回避问题。令人欣慰的是，这种“偷懒”的看法在暴胀宇宙论出现后受到了沉重的打击。因为按照暴胀宇宙论，宇宙创生之初即便存在正反物质的不对称，也会在暴胀过程中被稀释得微乎其微。因此初始条件并不能对今天观测到的正反物质的不对称给出令人满意的解释。

既然初始条件不足以解释正反物质的不对称，那我们就只能寄希望于宇宙创生之后所发生的具体物理过程了，这就是第二类看法。这类看法认为我们今天观测到的正反物质的不对称是由某些特定类型的物理过程产生的。

那么，究竟什么样的物理过程才能造成正反物质的不对称呢？早在1967年，苏联氢弹之父萨哈洛夫就提出了那样的物理过程所需满足的三个条件：

- （1）必须破坏费米子数守恒；
- （2）必须破坏C和CP对称性；
- （3）必须破坏热平衡。

这些条件后来被称为萨哈洛夫条件（Sakharov conditions），是任何能够

产生正反物质不对称的物理过程或物理理论所必须满足的。

萨哈洛夫条件中的第一条提到的费米子是组成物质的基本粒子，比如电子、质子和中子（进一步细分的话，质子和中子是由夸克组成的，而夸克也是费米子）。所有费米子的费米子数都是正的，而反费米子的费米子数则是负的。如果宇宙中的正反物质完全对称，那么总费米子数将是零。由于我们的宇宙中普通物质远比反物质多，因此总费米子数是正的。任何物理过程或物理理论要想让宇宙从正反物质完全对称（从而总费米子数为零）的状态演化到如今这个费米子数为正的状态，就必须改变总费米子数，从而必须破坏费米子数守恒。

萨哈洛夫条件中的第二条提到的C和CP对称性分别是基本粒子层次上的正反粒子对称性及正反粒子与宇称联合对称性。其中正反粒子对称性要求将一个物理过程中的所有粒子替换成相应的反粒子时，过程发生的几率不变。正反粒子与宇称联合对称性则是指在上述替换的同时再将物理过程换成它的镜像（好比是透过一面反射镜去看它）时，过程发生的几率也不变。这两个对称性之所以必须被破坏，是因为否则的话，任何可以造成物质多于反物质的物理过程都会伴随一个与它同样可能的、造成反物质多于物质的过程（即上述替换过程），这样两类过程的效果将会相互抵消。

最后，萨哈洛夫条件中的第三条之所以必须满足，是因为否则的话，任何可以造成物质多于反物质的物理过程都将与处在热平衡的逆过程相互抵消。

这三个条件虽被称为萨哈洛夫条件，不过萨哈洛夫本人在其长度只有三页的短文中其实并未如此鲜明地表述过这三个条件，这些条件是后人依据他的思路所归纳及重新表述的。

在这三个条件的基础上，物理学家们提出了许多理论模型，试图对正反物质不对称的起源作出定量解释。这些模型从相对简单的电弱统一理论（它是粒子物理标准模型的一部分），到各种各样的大统一理论，以及标准模型的超对称推广，种类繁多、应有尽有。但迄今为止，它们各自都存在一定的缺陷，或是结果的数量级不对，或是求解的困难度太大、或是过于特设、或是过于任意，尚无一个令人满意。不过尽管如此，现代物理为正反物质的不对称找到一个合理解释的前景看来是并不悲观的。

七、结语

我们有关反物质的介绍到这里就要结束了，虽然自人类发现反粒子迄今已有大半个世纪，但在理解物质与反物质的关系上还存在许多待解之谜。除了宇宙学尺度上正反物质的不对称外，在微观尺度上正反粒子也存在着令人困惑的不对称。物理学家们曾经认为，如果我们把一个微观物理过程中的所有粒子都替换成相应的反粒子，并且透过一面镜子去看它，那么我们所看到的新过程将与原过程有着相同的发生几率。这种对称性就是我们介绍萨哈洛夫条件时提到的CP对称性。由于这种对称性，反物质有时也被称为镜像物质。但令人困惑的是，这一对称性既非完全成立，也非完全不成立，而是非常接近成立^[6]。大自然为什么要让这面特殊的镜子如此接近完美却又不让它真正完美呢？我们不知道。

反物质是宇宙中的稀客，但这稀客是从相对意义上讲的，宇宙中反物质的绝对数量依然是极其庞大的，足以为科幻小说留下巨大的驰骋空间，这是值得庆幸的。只不过，反物质星球的存在看来是极不可能的，因为没有任何天然的物理过程能够让反物质有效地汇集起来，并在这一过程中免遭普通物质的“致命骚扰”。而反物质生物的存在则比反物质星球更加不可能得多，因为即便存在反物质星球，在那种星球上要想演化出生物来也是难以想象的。我们知道，即便在距离太阳系的形成已有约50亿年、太阳系空间已相当“干净”的今天，地球每天仍会受到上千万次的陨石撞击（这些陨石绝大多数在大气层中烧毁，只有少数落到地上，因此我们不必担心它们会恰好砸在我们头上），这些陨石的总质量约有几吨。这样的质量相对于庞大的地球来说无疑是微乎其微的，但同样的情形如果发生在一颗反物质星球上，那么这几吨的陨石（普通物质）与星球上的反物质湮灭所释放的能量将相当于上百万颗广岛原子弹爆炸所

释放的能量^[7]。要在一个每天被上百万颗原子弹轰击的地球上产生生物，这恐怕是最高级的想象力也难以胜任的。

因此，如果有朝一日我们与某种外星球的高等生物建立了联系，我们可以大大方方地伸出手去和他们相握（如果握手对他们来说也代表友善的话），而不必担心大家会在这样的亲密接触中相互湮灭^[8]。

2007年5月4日写于纽约

2014年10月23日最新修订

[1]比狄拉克稍早，瑞典物理学家克莱因（Oskar Klein）、德国物理学家高登（Walter Gordon）及奥地利物理学家薛定谔（Erwin Schrödinger）也提出了一个试图融合量子力学与相对论要求的方程：克莱因-高登方程（Klein-Gordon equation）。但克莱因-高登方程具有一些当时看来比狄拉克方程更令人不易接受的特征，延后了它被真正重视的时间。

[2]其实在经典相对论力学中也存在负能量状态，但在经典情况下我们可以摒弃负能量状态而不用担心它们对正能量状态产生影响，因为这两者之间存在一个非零的间隙（请读者想一想，对电子来说这一间隙有多大），而经典的物理过程都是连续的，从而不可能跨越这一间隙。

[3]正反粒子的湮灭产物可以是多种多样的。一般来说，参与湮灭的正反粒子的质量越大、能量越高，湮灭产物的种类通常就越多，在低能湮灭——尤其是轻粒子的低能湮灭——过程中，则有很大的几率产生光子对。

[4]在云室中比较同一种带电粒子的速度快慢是十分容易的，因为速度慢的粒子比速度快的粒子更容易被磁场所偏转，因此通过比较粒子径迹的偏转幅度——确切说是曲率——就可以比较出它们的速度快慢。

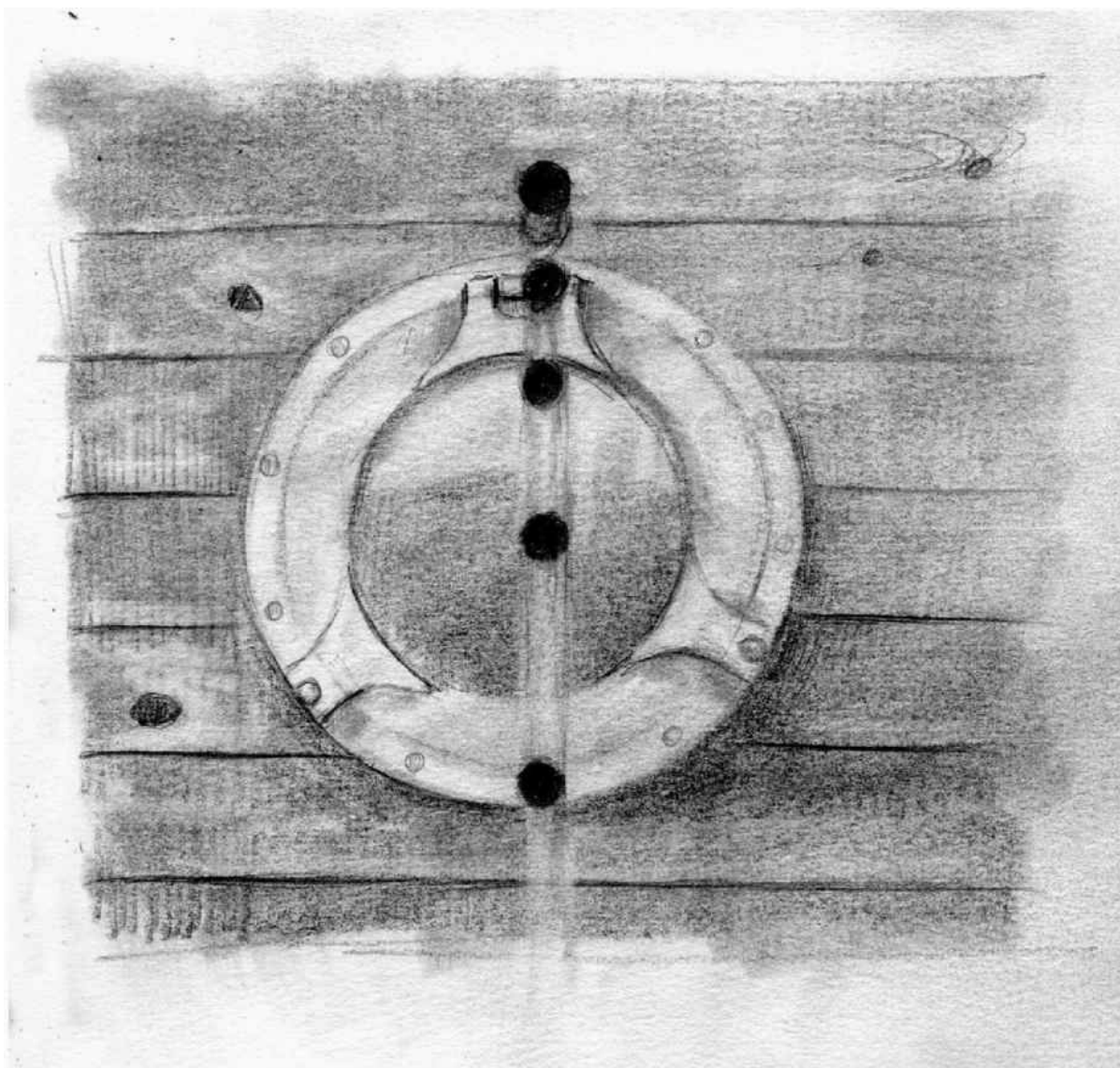
[5]值得一提的是，当时和安德逊一同在加州理工大学跟随密立根从事实验物理研究的中国物理

学家赵忠尧早在1929年至1930年间，就在研究硬 γ 射线穿越物质时，观测到了后来被证实为是源于正负电子对的产生的反常吸收效应，以及源于正负电子对的湮灭的特殊辐射——虽然这些实验并未直接观测正电子。

[6]在1957年以前，物理学家们想当然地认为所有这类离散对称性都是严格的，直到1957年宇称对称性倒下之后，才开始对离散对称性进行区分，但它们大都像多米诺骨牌似地也倒下了。CP是倒得比较慢的一个，前后也只经过了7年。

[7]有读者可能会问：为什么不干脆假定那些陨石也是反物质？从纯粹假定的角度上讲，自然是可以的，但我们的讨论有一个前提，那就是承认我们这个宇宙——如目前的理论与观测所表明的——是一个物质为主的宇宙。在这样的宇宙中，越大尺度的反物质分布就越不可能。因此我们对反物质出现的尺度只做最低限度的假定。

[8]不过，如果我们真的担心他们有可能是反物质构成的，也有办法在见面之前加以确认，确认的方法就是利用刚刚提到过的微观世界正反粒子之间的不对称性。李政道在其教材*Particle Physics and Introduction to Field Theory*（科学出版社出过中文版：《粒子物理和场论简引》）的第9.2节中对这一问题作了饶有趣味的论述，感兴趣的读者可以参阅。



绘画：张京

从伽利略船舱到光子马拉松^[1]

一、从相对性原理到相对论

现代人都知道，我们脚下的大地并不是静止不动的。事实上，在读者们阅读本文标题的短短一秒钟的时间里，我们脚下的大地已随着地球的自转移动了几百米（除非你很靠近两极），随着地球绕太阳的公转移动了约30千米，随着太阳系绕银河系中心的公转移动了约220千米。而我们的银河系也没闲着，它相对于所谓的宇宙微波背景辐射参照系移动了约550千米^[2]。这些运动大多数比火箭还快得多，人们却在很长的时间里一无所知，这是为什么呢？这个问题是我们的前辈在接受地球运动这一观念时面临的一大困扰，也是近代科学的一个启蒙性的问题。

近代科学的先驱者之一，意大利物理学家伽利略（Galileo Galilei）在名著《关于两大世界体系的对话》（*Dialogue concerning the Two Chief World Systems*）中对这一问题作了精彩的分析。伽利略注意到，地球运动的观念初看起来有违经验，其实却不然。相反，我们的经验表明，在一间封闭的船舱里，哪怕船在运动，只要运动得足够均匀，我们就无法发现它与处于静止时的任何区别。如果我们扔一块石头，往船头和船尾可以扔得一样远；如果我们观察一只小鸟的飞翔，它往哪个方向飞也都一样轻松。

我们现在知道，伽利略所注意到并归纳出的这一结果——即在所有

匀速运动的参照系中，自然现象由相同的规律所支配——是一条非常重要的物理学原理：相对性原理（principle of relativity）。不过在伽利略之后两百多年的时间里，物理学的发展虽然迅速，相对性原理却不曾有机会展示它的真正威力。

但是到了19世纪末，情况有了变化。那时候，物理学家们遇到了一个恼人的问题，那就是当时最成熟的两类物理学规律——力学和电磁学规律——似乎不能同时满足相对性原理。或者换句话说，如果力学规律满足相对性原理，那么电磁学规律就不满足相对性原理，反过来也一样。这个“鱼和熊掌”的局面令人深感为难，考虑到力学规律满足相对性原理是自伽利略以来就被牢固确立的事情，物理学家们大都决定舍电磁学而取力学。但问题是：舍电磁学意味着电磁学规律不满足相对性原理，从而也就意味着我们能通过在伽利略船舱里做某些电磁学实验，来分辨轮船的运动。

情况果真如此吗？

还别说，物理学家们真的做了那样的实验，他们选择了一条很特殊的大船：地球。毫无疑问，这是一条运动的大船，这一点在19世纪末已是凡地球人都知道的常识了。物理学家们所做的实验是什么呢？是一个测定电磁波速度的实验。如果电磁学规律不满足相对性原理，那么电磁波沿不同方向的传播速度就会不一样——除非地球恰好是静止的。实验的结果是什么呢？让人大跌眼镜，地球竟然真的是静止的！这下麻烦大了，难道兜了几个世纪的大圈子，我们又要重回地心说的年代？

幸运的是，这时有位名叫爱因斯坦（Albert Einstein）的专利局职员及时作出了一个相反的选择：舍力学而取电磁学。这样一来，所有证明地球静止的电磁学实验就都不再有效，比方说测定电磁波速度的实验就

会像在伽利略船舱中扔石头一样的无效。而我们——谢天谢地——也就不必重回地心说的年代了。但问题是：既然舍了力学，那力学规律该怎么办？爱因斯坦的回答很简单，那就是“削足适履”。既然力学规律这只脚放不进与电磁学规律相一致的相对性原理那只鞋，那就修改力学规律。

修改力学规律的结果是导致了一些很新奇的结果，比方说物体的质量原本被认为是常数，修改之后却变成与相对运动有关的了。爱因斯坦的这一回答实际上是把相对性原理提升为了一条比像力学、电磁学那样具体领域的物理理论都更基本的原理，由此建立的理论就是所谓的相对论（theory of relativity）。相对论在更广阔背景下再次确立了伽利略的观察，即在伽利略船舱中所做的任何实验或观测，都不可能分辨轮船的运动。

在此后一个多世纪的时间里，得到无数实验验证的相对论成为了现代物理学最坚实的基石之一。我们描述基本粒子的理论被称为相对论量子场论（relativistic quantum field theory），我们描述宇宙的理论被称为广义相对论（general theory of relativity）^[3]，我们描述日常现象的力学、电磁学等也全都满足相对论的要求。而当年的专利局职员则成为了有史以来最伟大的科学家之一。

一切似已尘埃落定。

但是，物理学家们注定是一群不安分守己的人，新的探索无论对于他们的好奇心还是职业都是必不可少的。相对论无疑是一座巍峨的高山，但物理学家们仍然要问：山的那边还有没有风景？

二、破坏相对论的思路与后果

物理学家们之所以要这样问，当然也有具体的原因。比方说我们前面提到的两个理论——描述基本粒子的相对论量子场论与描述宇宙的广义相对论——虽然各自都很成功，却迄今无法和睦共处。更糟糕的是，作为物理理论，它们又不可能做到井水不犯河水。因为在有些场合——比如在大质量、高密度的天体附近——哪怕是基本粒子之间的相互作用，也必须考虑引力的影响；又比如在宇宙大爆炸的初期，整个宇宙都处在微观尺度上，哪怕是最宏观的性质，也不能忽略量子效应。因此，相对论量子场论与广义相对论必须以某种方式融合到一起，这种融合是现代物理学所面临的最棘手的课题之一。

有意思的是，试图将这两个同时满足相对论要求的理论融合到一起的努力，却为破坏相对论的可能性开启了思路。

其中有一种努力的途径是认为问题的根源在于时空貌似光滑，其实却不然。当我们探索到只有原子核的一万亿亿分之一（ 10^{-20} ）的尺度——被称为普朗克尺度——上时，时空也许会显示出像网格一样的结构。这就好比一片丝绸，远远看去很光滑，拿到放大镜下，却可以看到密密层层网格结构。如果时空真的有那样的网格结构，那么伽利略船舱中的人只要有足够厉害的“放大镜”，就有可能通过观测时空的网格结构，来判断轮船是否在运动，从而破坏相对论的要求。

另一种努力的途径则是认为，时空中有可能存在一种被称为“背景场”的东西。这种东西不是由物质产生的，却能对物质施加影响（用物理学家们的术语来说，这是一种非动力学场），而且这种影响在不同位置、不同时刻，甚至对不同观测者都有可能是不一样的。如果说时空网

格像一片丝绸，那么这种背景场就像一种流体——比如水。在水中，即便我们无法像观察丝绸网格那样观察水分子，也依然可以判断物体的运动，因为我们可以观察水对物体的阻力。如果时空中真的存在那样的背景场，那么伽利略船舱中的人就可以通过观察它对普通物体的作用来判断轮船是否在运动，这同样破坏相对论的要求。

上面这些思路并非单纯的幻想，而是多少有一些物理上的缘由，甚至是某些理论模型的推论。比如时空的网格结构与一种被称为“圈量子引力”（loop quantum gravity）的理论不无渊源，而背景场的思路则可以从所谓的“超弦理论”（superstring theory）中获得某种支持^[4]。

破坏相对论这个潘多拉盒子一经打开，其他可能性也就应运而生了。比如有一种思路是这样的：将现实世界的物质全都扔掉，直接对相对论的数学结构开刀，由此可以得到一种被称为“双重狭义相对论”（doubly special relativity, DSR）的理论。这是一种很大胆的思路，可惜的是，迄今还没人知道如何将扔掉的物质重新放回到理论中去，因此这种思路的物理意义起码在目前还是成问题的^[5]。不过在一个连相对论都被怀疑的研究方向上，谁又敢说这种思路一定就没有可能呢？历史上纯粹源自数学考虑，却最终获得物理意义的例子毕竟还是有的，因此这样的思路也有一些人在研究。

看来破坏相对论的思路不仅有，而且还不止一条。

既然如此，那就让我们姑且假定相对论果真被破坏了。接下来的一个很重要的问题是：这种破坏会有什么后果？对这个问题的具体答案显然跟破坏相对论的具体方式有关，不过，由于破坏相对论的思路大都与时空的结构有关，而时空是引力的源泉，因此我们可以预期，破坏相对论的后果之一，就是使引力发生变化。

比方说，如果破坏相对论的肇事者是背景场，就有可能对引力产生影响。我们在前面提到过，背景场能对物质施加影响，这种影响的可能的体现方式之一就是对引力的修正。而且这种修正在不同位置、不同时刻可以是不同的——或者用一些科普报道所用的比喻来说，是苹果在不同季节的掉落快慢有可能是不同的。

除了苹果的掉落快慢有可能不同这样的“家常”后果外，破坏相对论还可能造成一些更严重的后果。比方说，相对论中有一条很基本的原理，叫做光速不变原理^[6]，它表明光速是一个普适的极限速度。在很多破坏相对论的理论中，这条原理不再成立，不同的粒子可以有不同的极限速度。初看起来，这似乎没什么大不了，但是有科学家研究后发现，利用这一结果可以在黑洞附近让热量自发地从低温物体传向高温物体^[7]。这是一个令人吃惊的结果，因为在自然界中，热量的自发传输一向是从高温物体传向低温物体，而不能相反。这是一条很重要的物理学原理，叫做热力学第二定律，违反这一原理的物理过程被称为第二类永动机，它与违反能量守恒定律的第一类永动机一样，被认为是不可能实现的。

因此，破坏相对论的后果很可能是牵一发而动全身的，它所引发多米诺骨牌效应，很可能导致其他一些很重要的物理学原理也被破坏。这其实是可以预期的，因为物理学是一个整体，它的各个分支之间有着千丝万缕的关联，它的基础并不是一系列孤立假设的集合，我们很难在破坏像相对论那样的重要部分时不影响到其他部分。

三、光子的马拉松——破坏相对论的证据？

以上我们介绍了很多理论上的东西，在物理学上，再雄辩的理论也离不开观测与实验的评判。对于相对论的破坏来说，它即便存在也极其微弱，我们该如何去寻找观测与实验的评判呢？在当前的条件下，比较希望的探索方向主要有两类。

一类是探索微观世界的对称性破缺。这类探索有一段不短的历史。在1957年以前，人们曾经以为微观世界充满了对称性，其中很重要的一条是说微观世界的规律可以通过一面镜子去看而不被改变——这被称为宇称（parity）对称性。可惜这一对称性在1957年被证实是破缺的——确切地说是在所谓弱相互作用中是破缺的。不过这一对称性还可以加强，比如在通过镜子去看的同时把粒子与反粒子对换，可惜就连这种加强版的对称性在1964也被证实是破缺的——也是在所谓弱相互作用中破缺。但这一对称性还有一个终极加强版，那就是在通过镜子去看的同时，不仅把粒子与反粒子对换，而且让时间倒流。一些理论研究表明，在某些合理的条件下，这种终极加强版的对称性与相对论几乎是“一条绳上的两只蚂蚱”，一旦前者遭到破坏，后者也难以独善其身^[8]。按照这一结果，只要我们能在微观世界里找到任何确凿的现象破坏这种终极加强版的对称性——比如发现任何一个基本粒子的质量、自旋、电荷、衰变方式等性质与反粒子不严格对应——就相当于间接证实了相对论的破坏。这方面的实验数据可以说是天天都在积累（虽然目的大都不是为了证实相对论的破坏），但迄今尚无任何证据显示相对论被破坏。

另一类探索在思路更为直接。我们刚才提到过，在很多破坏相对论的理论中，光速不变原理不再成立。由此导致的结果，是不同的粒子

可以有不同的极限速度。但除此之外，它往往还意味着不同能量的光子在真空中的传播速度彼此不同——这被称为真空色散（vacuum dispersion）。利用这一特点，我们可以让不同能量的光子进行跑步比赛，来观察它们的速度是否不同，进而判断相对论是否被破坏^[9]。不过由于光子的速度实在太快，彼此的速度差异又即便有也极其细微，要想分出胜负，比赛必须是马拉松，而赛场只能是星空。

2005年夏天，天文学家们终于观察到了这样一次马拉松，一群高能光子从一个编号为“马卡良501”（Markarian 501）的遥远的活动星系核出发，经过5亿年的漫长旅程，抵达了地球。这群光子是一次伽马射线耀斑（gamma ray flare）的产物，它们的抵达被位于西班牙西南加那利群岛（Canary islands）上的“大气伽马切伦科夫成像望远镜”（major atmospheric gamma-ray imaging Cherenkov telescope, MAGIC）所记录。在记录中令科学界感到震动的是，能量在1.2~10TeV之间的高能光子的到达时间比能量在0.25~0.6TeV之间的低能光子晚了约4分钟，这与某些破坏相对论的理论所预期的大致相符。

那么，我们是不是可以就此宣布相对论被破坏了呢？不能。因为我们对这场5亿年前就起跑的马拉松知道得还太少，高能光子的到达时间虽然晚了4分钟，但它的起跑是否也晚了呢？我们却一无所知。

而更有意思的是，2009年，科学家们通过翱翔在外层空间的“费米伽马射线太空望远镜”（Fermi gamma-ray space telescope, FGST）又观测到了一次光子马拉松（图6）。参加这次马拉松的光子来自一次伽马射线暴（gamma ray burst），它的威力比产生前一次马拉松的伽马射线耀斑还要巨大得多，距离也更遥远得多（红移值约为0.9）。那些光子经过了数量级为百亿年的漫长跋涉才抵达地球，这几乎是我们这个宇宙所能提供的最长的赛程。这赛程是如此之长，以至于在这次马拉松起跑

的时候，不仅我们不存在，就连我们脚下这颗蓝色星球都尚未形成！与上次不同的是，这次马拉松的结果是高能光子（能量约为 31GeV ）与低能光子（能量在 10keV 以下）几乎同时到达终点（时间差在几十毫秒到几秒之间，几乎可以忽略），从而不仅没有破坏相对论，反而几乎给所有破坏相对论的理论下达了死亡通知书^[10]。



图6 费米伽马射线太空望远镜

两次光子马拉松，一对彼此相反的结果，我们究竟该相信什么呢？答案恐怕是：什么都先别相信，去寻找更多的证据。著名的美国行星天文学家萨根（Carl Sagan）有一句名言：超常的主张需要超常的证据（extraordinary claims require extraordinary evidence）^[11]。在相对论所具有的庞大的证据链面前，破坏相对论的理论无疑是超常的主张，但那两次光子马拉松却绝非超常的证据（更不用说它们还彼此矛盾），对所有有志于这一领域的研究者来说，探索的路还很漫长。

2009年9月25日写于纽约

2014年11月9日最新修订

^[1]本文是应《科学画报》约稿而写的关于破坏相对论的可能性的科普，原本有几段文字针对的是编辑指定的《新科学家》（*New Scientist*）杂志所报道的一个新理论，但由于该理论在所介绍

的领域内并无特殊重要性，修订时我删去了与该理论有关的内容，使本文成为了一般性的介绍。

[2]由于这些运动的方向各不相同，因此地球相对于宇宙微波背景辐射参照系的运动并不是上述数字的简单相加，而必须考虑方向的因素。观测表明，太阳系相对于宇宙微波背景辐射参照系的运动速度约为每秒370千米。（请读者想一想，我们为什么不给出地球的运动速度？）

[3]在广义相对论的每个时空点附近足够小的区域内，都可以找到特殊的参照系，在其中物理规律与在匀速运动的参照系中一样，这就好比光滑曲面上每个点附近足够小的区域都很接近平面一样。

[4]超弦理论本身是符合相对论要求的——确切地说是具有洛伦兹对称性（Lorentz symmetry）的，超弦理论中的相对论破坏（确切地说是指破坏洛伦兹对称性）是以对称性自发破缺的形式出现的。

[5]具体地说，双重狭义相对论是通过对动量空间中的庞加莱代数（Poincaré algebra）进行修改而来的，因此有时也被称为变形狭义相对论（deformed special relativity，缩写恰好仍是DSR）。双重狭义相对论除了像狭义相对论一样存在一个不变速度外，还存在一个不变动量（名称中的“双重”一词便由此而来）。双重狭义相对论的部分特点可以在某些非对易几何模型中找到渊源（但也只是数学渊源），另有些人则希望（目前还只是奢望）它能与圈量子引力建立联系。但迄今为止，该理论只有运动学，而无动力学，甚至连自洽性都尚待澄清。

[6]这条原理是让电磁学规律满足相对性原理的必然推论。

[7]这是2006年俄罗斯科学院核子研究所（Institute for Nuclear Research of the Russian Academy of Sciences）的两位物理学家在《物理快报》（*Physics Letters*）上发表的一个结果。他们的大致思路是这样的：不同的粒子具有不同的速度上限意味着黑洞辐射中不同的粒子会有不同的辐射温度。假定粒子B的辐射温度高于粒子A，我们在黑洞外面构筑两个壳层，壳层A只能发射和吸收粒子A，壳层B只能发射和吸收粒子B，我们选择壳层的温度使得（粒子B的辐射温度）>（壳层B的温度）>（壳层A的温度）>（粒子A的辐射温度）。在这样的安排下，壳层A会通过粒子A

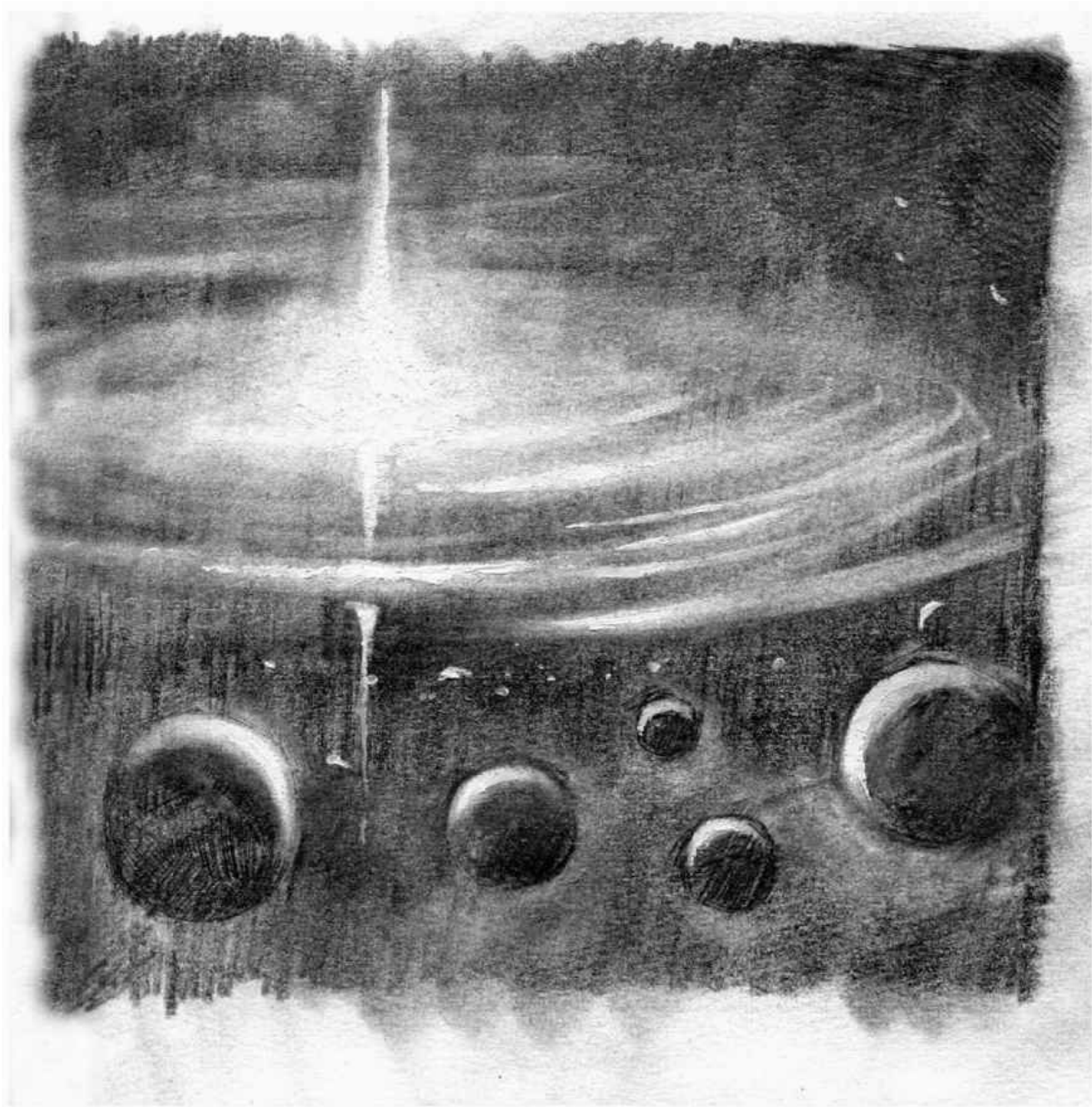
将热量传给黑洞，而黑洞又会通过粒子B将热量传给壳层B，净效果是壳层A将热量传给壳层B，即热量自发地从低温物体传往了高温物体。

[8]但反过来则不然，即相对论的破坏不一定意味着那种终极加强版的对称性——即所谓的CPT对称性——的破坏。因此严格地讲，它们并不完全是“一条绳上的两只蚂蚱”。

[9]确切地讲，不同能量的光子具有相同速度可以推翻许多破坏相对论的理论，但相反的结果，即不同能量的光子具有不同速度，却并不能直接证实相对论的破坏，因为相对论所要求的只是存在一个不变速度，这个速度不一定非得是光子的速度，甚至不一定非得有任何粒子具有这一不变速度。

[10]因为如果这次光子马拉松的结果可信，那么破坏相对论的效应将会细微到不自然的程度，比方说对于最简单的真空色散模型——即色散率的修正项线性正比于能量的模型——来说，破坏相对论的能标将会比所谓的“普朗克能标”（Planck scale）还高得多。

[11]萨根的这一表述具有较大的公众影响，不过他并不是最早提出这类原则的人，早在两百多年前，法国数学家拉普拉斯（Pierre-Simon Laplace）就曾说过：“支持一个超常主张的证据分量必须正比于主张的奇异程度。”



绘画：张京

质量的起源^[1]

一、引言

物理学是一门试图在最基本的层次上理解自然的古老科学，它的早期曾经是哲学的一部分。在那个时期，物理学所关心的是一些有关世界本原的问题。那些问题看似朴素，却极为困难。在后来的漫长岁月里，物理学曾经一次次地回到那些问题上来，就像远行的水手一次次地回望灯塔。

“质量的起源”便是一个有关世界本原的问题。

二、宇宙物质的组成

我们首先来界定一下所要讨论的质量究竟是什么东西的质量。这在以前是不言而喻的，现在的情况却有了变化，因此有必要加以界定。众所周知，过去十年里观测宇宙学所取得的一个令人瞩目的成就，就是以较高的精度测定了宇宙物质的组成，从而使我们在宇宙学的历史上第一次可以谈论所谓的“精密宇宙学”（precision cosmology）。

按照这种“精密宇宙学”为我们绘出的图景，在宇宙目前的能量密度中暗能量（dark energy）约占68%，暗物质（dark matter）约占27%，而我们熟悉的所谓“可见物质”（visible matter）或“普通物质”（ordinary matter）只占可怜兮兮的5%。在这些组成部分中，对暗能量与暗物质的研究目前还处于很初级的阶段，尚未建立起足够具体且有实验基础的理论。因此本文对之不做讨论。

除去了暗能量与暗物质，剩下的就是可见物质了。可见物质在宇宙能量密度中所占的比例虽小，却是我们所熟知的物质世界的主体。可观测宇宙中数以千亿计的星系，每个星系中数以千亿计的恒星，以及某个不起眼的恒星附近第三颗行星上数十亿的灵长类生物，全都包含在了这小小的5%的可见物质之中^[2]。

本文要讨论的便是这可见物质。

与“暗”字打头的其余95%的能量密度相比，我们对可见物质的研究与了解无疑要深入得多。今天几乎每一位中学生都知道，这部分物质主要是由质子、中子、电子等粒子组成的。因此很明显，要讨论质量的起源，归根到底是要讨论这些粒子的质量起源。

三、从机械观到电磁观

对几乎所有受过现代教育的人来说，最早接触质量这一物理概念都是在牛顿力学中。在牛顿力学中，质量是决定物体惯性和引力的基本物理量，是一个不可约（irreducible）的概念。我们知道，在大约两百年的时间里，牛顿力学被认为是描述物理世界的基本框架，这就是所谓的机械观（mechanical worldview）。在那段时间里，物理学家们曾经试图把物理学的各个分支尽可能地约化为力学。很显然，在那样一个以机械观为主导的时期里，质量既然是力学中的不可约概念，自然也就成为了整个物理学中的不可约概念。不可约概念顾名思义，就是不需要也不能够约化为更基本的概念的，因此有关质量起源的研究在那个时期是基本不存在的^[3]。

但是到了19世纪末的时候，试图把物理学的各个分支约化为力学的努力遭到了很大的挫折。这种挫折首先来自于电磁理论。大家知道，电磁理论预言了电磁波。按照机械观，波的传播必然有相应的介质。但电磁波是在什么介质中传播的呢？却是谁也不知道。尽管如此，物理学家们还是按照机械观的思路假设了这种介质的存在，并称之为“以太”（aether）。但不幸的是，所有试图为以太构筑机械模型的努力全都在实验面前遭遇了滑铁卢。在那段最终催生了狭义相对论的物理学阵痛期里，许多物理学家艰难地试图调和着实验与机械以太模型之间的矛盾。但与那些挽救机械观的努力同时，一种与机械观截然相反的思路也萌发了起来，那便是电磁观（electromagnetic worldview）。电磁观的思路是：物理学上并没有什么先验的理由要求我们用力学的框架来描述自然，机械观的产生只不过是力学在很长一个时期里是发展最为成熟的物理学分支而已，现在电磁理论也发展到了不亚于力学的成熟程度，

既然无法把电磁理论约化为力学，那何不反过来把力学约化为电磁理论呢？

要想把力学约化为电磁理论，一个很关键的步骤就是把力学中的不可约概念——质量——约化为电磁概念，这是物理学家们研究质量起源的第一种定量尝试。由于当时对物质的微观结构还知之甚少，1897年由汤姆逊（Joseph John Thomson, 1856—1940年）所发现的电子是当时所知的唯一的基本粒子，因此将质量约化为电磁概念的努力就集中体现在了对电子的研究上，由此产生了物理史上昙花一现的经典电子论（classical electron theory）。

四、经典电子论

经典电子论最著名的人物是荷兰物理学家洛伦兹（Hendrik Lorentz, 1853—1928年），他是一位经典物理学的大师。在相对论诞生之前的那几年里，洛伦兹虽已年届半百，却依然才思敏捷。1904年，洛伦兹发表了一篇题为《任意亚光速运动系统中的电磁现象》（*Electromagnetic Phenomena in a System Moving with Any Velocity Less than that of Light*）的文章。在这篇文章中他运用自己此前几年在研究运动系统的电磁理论时所提出的包括长度收缩（length contraction）、局域时间（local time）在内的一系列假设，计算了具有均匀面电荷分布的运动电子的电磁动量，由此得到电子的横质量 m_T 与纵质量 m_L 分别为（这里用的是高斯单位制^[4]）：

$$m_T = \frac{2}{3} \frac{e^2}{Rc^2} \gamma, \quad m_L = \frac{2}{3} \frac{e^2}{Rc^2} \gamma^3$$

其中 e 为电子的电荷， R 为电子在静止参照系中的半径， c 为光速， $\gamma = (1 - v^2/c^2)^{-1/2}$ 。撇开系数不论，洛伦兹这两个结果所包含的质量与速度的关系与后来的狭义相对论完全相同。

但洛伦兹的文章刚一发表就遭到了经典电子论的另一位主要人物亚伯拉罕（Max Abraham, 1875—1922年）的批评。亚伯拉罕指出，质量除了像洛伦兹那样通过动量来定义，还应该可以通过能量来定义。比方说纵质量可以定义为 $m_L = (1/v) (dE/dv)$ ^[5]。但简单的计算表明，用这种方法得到的质量与洛伦兹的结果完全不同。

这说明洛伦兹的电子论是有缺陷的。那么缺陷在哪里呢？亚伯拉罕

认为是洛伦兹的计算忽略了为平衡电子内部各电荷元之间的相互排斥所必需的张力。没有那样的张力，洛伦兹的电子会在各电荷元的相互排斥下土崩瓦解^[6]。除亚伯拉罕外，另一位经典物理学大师庞加莱（Henri Poincaré, 1854—1912年）也注意到了洛伦兹电子论的这一问题。庞加莱与洛伦兹是爱因斯坦之前在定量结果上最接近狭义相对论的物理学家。不过比较而言，洛伦兹的工作更为直接，为了调和以太理论与实验的矛盾，他提出了许多具体的假设，而庞加莱往往是在从美学与哲学角度审视洛伦兹及其他人的工作时对那些工作进行修饰及完善。这也很符合这两人的特点，洛伦兹是一位第一流的工作型物理学家（working physicist），而庞加莱既是第一流的数学及物理学家，又是第一流的科学哲学家。在1904年至1906年间，庞加莱亲自对洛伦兹电子论进行了研究，并定量地引进了为维持电荷平衡所需的张力，这种张力因此而被称为庞加莱张力（Poincaré stress）。在庞加莱工作的基础上，1911年，即在爱因斯坦与闵科夫斯基（Hermann Minkowski, 1864—1909年）建立了狭义相对论的数学框架之后，德国物理学家冯·劳厄（Max von Laue, 1879—1960年）证明了带有庞加莱张力的电子的能量动量具有正确的洛伦兹变换规律。

下面我们用现代语言来简单叙述一下经典电子论有关电子结构的这些主要结果。按照狭义相对论中最常用的约定，我们引进两个惯性参照系： S 与 S' ， S' 相对于 S 沿 x 轴以速度 v 运动。假定电子在 S 系中静止，则在 S' 系中电子的动量为

$$p'^{\mu} = \int_{t'-0} T'^{0\mu}(x'^{\xi}) d^3 x' = L_{\alpha}^0 L_{\beta}^{\mu} \int T^{\alpha\beta}(x^{\xi}) d^3 x'$$

其中 T 为电子的总能量动量张量， L 为洛伦兹变换矩阵。由于 S 系中 $T^{\alpha\beta}$ 与 t 无关，考虑到

$$\int T^{\alpha\beta}(x^\xi) d^3x' = \int T^{\alpha\beta}(\gamma x', y', z') d^3x' = \gamma^{-1} \int T^{\alpha\beta}(x^\xi) d^3x$$

上式可改写为

$$p'^\mu = \gamma^{-1} L_\alpha^0 L_\beta^\mu \int T^{\alpha\beta}(x^\xi) d^3x$$

由此得到电子的能量与动量分别为（有兴趣的读者可试着自行证明一下）

$$E = p'^0 = \gamma m + \gamma^{-1} L_i^0 L_j^0 \int T^{ij}(x^\xi) d^3x$$

$$p = p'^1 = \gamma m v + \gamma^{-1} L_i^0 L_j^1 \int T^{ij}(x^\xi) d^3x$$

这里 i, j 的取值范围为空间指标1, 2, 3, $m = \int T^{00}(x^\xi) d^3x$, 为了简化结果, 我们取 $c=1$ 。显然, 由这两个式子的第一项所给出的能量动量是狭义相对论所需要的, 而洛伦兹电子论的问题就在于当 $T^{\mu\nu}$ 只包含纯电磁能量动量张量 $T_{\text{EM}}^{\mu\nu}$ 时这两个式子的第二项非零^[7]。

那么庞加莱张力为什么能避免洛伦兹电子论的这一问题呢？关键在于引进庞加莱张力后电子才成为一个满足力密度 $f^\mu = \partial_\nu T^{\mu\nu} = 0$ 的孤立平衡体系。在电子静止系 S 中 $T^{\mu\nu}$ 不含时间, 因此 $\partial_j T^{ij} = 0$ 。由此可以得到一个很有用的关系式（请读者自行证明）： $\partial_k (T^{ik} x^j) = T^{ij}$ 。对这个式子做体积分, 注意到左边的积分为零, 便可得到

$$\int T^{ij}(x^\xi) d^3x = 0$$

这个结果被称为冯·劳厄定理（von Laue's theorem），它表明我们上面给出的电子能量动量表达式中的第二项为零。因此庞加莱张力的引进非常

漂亮地保证了电子能量动量的协变性。

至此，经过洛伦兹，庞加莱，冯·劳厄等人的工作，经典电子论似乎达到了一个颇为优美的境界，既维持了电子的稳定性，又满足了能量动量的协变性。但事实上，在这一系列工作完成时经典电子论对电子结构的描述已经处在了一个看似完善，实则没落的境地。这其中的一个原因便是那个“非常漂亮地”保证了电子能量动量协变性的庞加莱张力。这个张力究竟是什么？我们几乎一无所知。更糟糕的是，若真的完全一无所知倒也罢了，我们却偏偏还知道一点，那就是庞加莱张力必须是非电磁起源的（因为它的作用是抗衡电磁相互作用），而这恰恰是对电磁观的一个沉重打击。

就这样，试图把质量约化为纯电磁概念的努力由于必须引进非电磁起源的庞加莱张力而化为了泡影。但这对于很快到来的经典电子论及电磁观的整体没落来说还只是一个很次要的原因。

五、量子电动力学

经典电子论的没落是物理学史上最富宿命色彩的事件。这一宿命的由来是因为电子发现得太晚，而量子理论又出现得太早，这就注定了夹在其间，因“电子”而始、逢“量子”而终的经典电子论只能有一个昙花一现的命运^[8]。为它陪葬而终还有建立在经典电磁理论基础上的整个电磁观。

量子理论对经典物理学的冲击是全方位的，足可写成一部壮丽的史诗。就经典电子论中有关电子结构的部分而言，对这种冲击最简单的启发性描述来自于所谓的不确定原理（uncertainty principle）。如我们在第四节中看到的，经典电子论给出的电子质量——除去一个与电荷分布有关的数量级为1的因子——约为 e^2/Rc^2 。由此可以很容易地估算出 $R \sim 10^{-15}$ 米（感兴趣的读者请自行验证一下）。这被称为电子的经典半径。但是从不确定原理的角度看，对电子的空间定位精度只能达到电子的康普顿波长 $h/mc \sim R/\alpha \sim 10^{-12}$ 米的量级（其中 $\alpha \approx 1/137$ 为精细结构常数），把电子视为经典电荷分布的做法只有在空间尺度远大于这一量级的情形下才适用。由于电子的经典半径远远小于这一尺度，这表明经典电子论并不适用于描述电子的结构。建立在经典电子论基础上的电子质量计算也因此而失去了理论基础^[9]。

但是经典电子论对电子质量的计算虽然随着量子理论的出现而丧失了理论基础，那种计算所体现的相互作用对电子质量具有贡献的思想却是合理的，并在量子理论中得到了保留。这种贡献被称为电子自能（electron self energy）。在量子理论基础上对电子自能的计算最早是由瑞典物理学家沃勒（Ivar Waller, 1898—1991年）于1930年在单电子狄拉

克理论的基础上给出的，结果随虚光子动量的平方而发散。1934年奥地利裔美国物理学家韦斯科夫（Victor Weisskopf, 1908—2002年）计算了狄拉克空穴理论（hole theory）下的电子自能，结果发现其发散速度比沃勒给出的慢得多，只随虚光子动量的对数而发散^[10]。撇开当时那些计算所具有的诸多缺陷不论，韦斯科夫的这一结果在定性上是与现代量子场论一致的。

按照现代量子场论，相互作用对电子自能的贡献可以用对电子传播子产生贡献的单粒子不可约图（one-particle irreducible diagrams）来描述，其中主要部分来自量子电动力学（Quantum Electrodynamics, QED）所描述的电磁自能，而电磁自能中最简单的贡献则来自

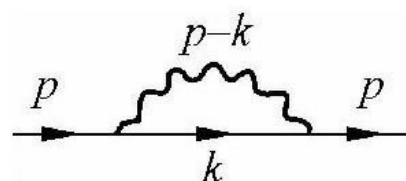


图7 最简单的电子自能图

于如图7所示的单圈图。幸运的是，由于量子电动力学的耦合常数在所有实验所及的能区都很小，因此这个最简单的单圈图的贡献在整个电子自能中占了主要部分^[11]。

对这一单圈图的计算在任何一本量子场论教材中都有详细介绍，其结果为 $\delta m \sim \alpha m \ln(\Lambda/m)$ ，其中 m 为出现在量子电动力学拉氏量中的电子质量参数，被称为裸质量（bare mass）， Λ 为虚光子动量的截断（cut-off）能标。如果我们把量子电动力学的适用范围无限外推，允许虚光子具有任意大的动量，则 δm 将趋于无穷，这便是自20世纪三四十年代起困扰物理学界几十年之久的量子场论发散困难的一个例子。

量子场论中的发散困难，究其根本是由所谓的点粒子模型引起的。这种发散具有相当的普遍性，不单单出现在量子场论中。将经典电子论运用于点电子模型同样会出现发散，这一点从经典电子论的电子质量公

式 $m \sim e^2/Rc^2$ 中可以清楚地看到：当电子半径 R 趋于零时质量 m 趋于无穷。经典电子论通过引进电子的有限半径（从而放弃点粒子模型）免除了这一发散，但伴随而来的庞加莱张力、电荷分布等概念却在很大程度上使电子丧失了基本粒子应有的简单性^[12]。这种简单性虽没有先验的理由，但毫无疑问是人们引进基本粒子这一概念时怀有的一种美学上的期待，正如狄拉克所说：“电子太简单，支配其结构的定律根本不应该成为问题。”经典电子论将质量约化为电磁概念的努力即便在其他方面都成功了，其意义也将由于引进电子半径这一额外参数及庞加莱张力、电荷分布等额外假设而大为失色。从这一角度上讲，量子电动力学在概念约化上比经典电子论显得更为彻底，因为在量子电动力学的拉氏量中不含有任何与基本粒子结构有关的几何参数。基本粒子在量子场论中是以点粒子的形式出现的，虽然这并不意味着它们不具有唯象意义上的等效结构，但所有那些结构都是作为理论的结果而不是如经典电子论中那样作为额外假设而出现的，这是除与狭义相对论及量子理论同时兼容，与实验高度相符之外，建立在点粒子模型基础上的量子场论又一个明显优于经典电子论的地方。

至于由此产生的发散困难，在20世纪70年代之后随着重整化（renormalization）方法的成熟而得到了较为系统的解决。不过尽管人们对重整化方法在数学计算及物理意义的理解上都已相当成熟，发散性的出现在很多物理学家眼里仍基本消除了传统量子场论成为所谓“终极理论”（theory of everything）的可能性，这是后话。

六、质量电磁起源的破灭

既然量子电动力学与经典电子论一样具有电子自能，那它能否代替经典电子论实现后者没能实现的把质量完全约化为电磁概念的梦想呢？很可惜，答案是否定的。

这可以从两方面看出来。

首先，从 $\delta m \sim \alpha m \ln(\Lambda/m)$ 中可以看到，由电磁自能产生的质量修正 δm 与裸质量 m 的比值为 $\alpha \ln(\Lambda/m)$ 。由于 $\alpha \approx 1/137$ 是一个比较小的数目， $\ln(\Lambda/m)$ 又是一个增长极其缓慢的函数，因此对于任何普朗克能标以下的截断， $\ln(\Lambda/m)$ 都是一个比较小的数目（特别是，这一数目小于1）。这意味着由电磁自能产生的质量修正是比较小的——比裸质量更小^[13]。

另一方面，即便我们一厢情愿地把量子电动力学的适用范围延伸到比普朗克能标还高得多的能区，从而使 δm 变得很大，把质量完全约化为电磁概念的梦想依然无法实现。因为电子的电磁自能还有一个很要命的特点，那就是 $\delta m \propto m$ 。这表明，无论把截断能标取得多大，如果裸质量为零，电子的电磁自能也将为零。因此，为了解释电子质量，裸质量不能为零，而裸质量作为量子电动力学拉氏量中的参数，在量子电动力学的范围之内是无法约化的，从而终结了在量子电动力学中把质量完全约化为电磁概念的梦想。

有的读者可能会问：电磁自能既然是由电磁相互作用引起的，理应只与电荷有关，为什么却会正比于裸质量呢？这其中的奥妙在于对称性。量子电动力学的拉氏量：

$$L = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - e\bar{\psi}\gamma^\mu A_\mu\psi$$

在 $m=0$ 时具有一种额外的对称性，即在 $\psi \rightarrow e^{i\alpha\gamma^5}\psi$ 下不变（请有兴趣的读者自行证明）。这种对称性被称为手征对称性（chiral symmetry），它表明在 $m=0$ 的情形下电子的左右手征态：

$$\psi_L = \frac{1-\gamma^5}{2}\psi, \quad \psi_R = \frac{1+\gamma^5}{2}\psi$$

不会互相耦合。另一方面，（读者可以很容易地证明）电子的质量项

$$m\bar{\psi}\psi = m\bar{\psi}_L\psi_R + m\bar{\psi}_R\psi_L$$

却是一个电子左右手征态相互耦合，从而破坏手征对称性的项。这样的项在电子的裸质量不存在——从而量子电动力学的拉氏量具有手征对称性——的情况下将被手征对称性所禁止，不可能出现在任何微扰修正中。因此 $\delta m \sim \alpha m \ln(\Lambda/m)$ 这一结果的出现是很自然的^[14]。

至此我们看到，试图把质量完全归因于电磁相互作用的想法在量子理论中彻底地破灭了。电磁质量即便在像电子这样质量最小——从某种意义上讲也最为纯粹——的带电粒子的质量中也只占一个不大的比例，在其他粒子——尤其是那些不带电荷的基本粒子——中就更甭提了。

很显然，质量的主要来源必须到别处去寻找。

七、对称性自发破缺

质量的电磁起源破灭后，质量起源问题沉寂了很长一段时间。但物理学本身的前进步伐并未因此而停顿。物理学家们手头有大量的观测数据需要分析和解释，同时理论体系本身也有大量的问题亟待解决。对现代物理学的发展来说，这些具体或细节问题是远比解决像质量起源那样的本原问题更重要的动力。另一方面，现代物理学在研究这些具体或细节问题中逐渐积累起来的智慧与洞见，又常常会为更深入地探求本原问题提供新的思路。这是现代物理学的卓越之处，也是它没有像那些只注重于深奥的本原问题，却对细节不屑一顾的其他尝试那样流于肤浅的重要原因。

物理学再次回到质量起源问题是在20世纪60年代。

在20世纪60年代初的时候，物理学家们在对基本粒子的研究中已经发现了许多对称性。对称性在物理学中一直有着重要地位，不仅由于其优美的形式与某些物理学家对自然规律的美学追求十分吻合，更重要的是因为它们不仅中看，而且中用，有一种穿透复杂性的力量。即便在对一个物理体系的动力学行为还缺乏透彻理解的情况下，对称性也往往具有令人瞩目的预言能力。这最后一点在20世纪五六十年代的粒子物理研究中具有极大的吸引力，因为当时人们对基本粒子相互作用的动力学机制还知之甚少，而且对在很大程度上为研究基本粒子相互作用而发展起来的量子场论产生了很深的怀疑。在这种情况下，许多物理学家对对称性寄予了厚望，希望通过它们来窥视大自然在这一层次上的奥秘。

但不幸的是，当时所发现的许多对称性却被证明只在近似的情况下才成立，比如同位旋对称性。如何理解这种近似的对称性呢？当时有一

种猜测，认为近似对称性是（严格）对称性自发破缺的产物。

所谓对称性自发破缺（spontaneous symmetry breaking），指的是这样一种情形：即一个物理体系的拉氏量具有某种对称性，而基态却不具有该对称性。换句话说，体系的基态破缺了运动方程所具有的对称性。这种对称性自发破缺的概念最早是出现在凝聚态物理中的，20世纪60年代被日裔美国物理学家南部阳一郎（Yoichiro Nambu, 1921—）和意大利物理学家约纳-拉西尼奥（Giovanni Jona-Lasinio, 1932—）引进到量子场论中。在量子场论中，体系的基态是真空态，因此对称性自发破缺表现为体系拉氏量所具有的对称性被真空态所破缺。

有的读者可能会问：一个物理体系的真空态是由拉氏量所确定的，为什么会不具有拉氏量所具有的对称性呢？这其中的奥秘在于许多物理体系具有简并的真空态，如果我们把所有这些简并的真空态视为一个集合，它的确与拉氏量具有同样的对称性。但物理体系的实际真空态只是该集合中的一个态，这个态往往不具有整个集合所具有的对称性，这就造成了对称性的破缺——也就是我们所说的对称性自发破缺^[15]。

但是把近似对称性归因于对称性自发破缺的想法在1961年遭到了致命的打击。那一年由英国物理学家戈德斯通（Jeffrey Goldstone, 1933—）提出并在稍后与巴基斯坦物理学家萨拉姆（Abdus Salam, 1926—1996年）及美国物理学家温伯格（Steven Weinberg, 1933—）一起证明了这样一个命题——被称为戈德斯通定理（Goldstone theorem）：每一个自发破缺的整体连续对称性都必然伴随一个无质量标量粒子。这个无质量标量粒子被称为戈德斯通粒子（Goldstone particle）或南部-戈德斯通粒子（Nambu-Goldstone particle）。

为什么会有这样的结果呢？我们来简单地证明一下：

假定一个物理体系的拉氏量中的势函数为 $V(\varphi_a)$ ($a=1, \dots, N$)，其中 φ_a 为标量场（可以是基本的也可以是复合的）。显然，该体系的真空态满足 $\partial V/\partial \varphi_a = 0$ （为避免符号繁复，我们略去了对真空的标记），而标量粒子的质量（平方）由 $\partial^2 V/\partial \varphi_a \partial \varphi_b$ 在真空态上的本征值给出。现在考虑对真空态 φ_a 作一个无穷小连续对称变换 $\varphi_a \rightarrow \varphi_a + \epsilon \Delta_a(\varphi)$ （其中 ϵ 为无穷小参数）。由于 $V(\varphi_a)$ 在这一变换下不变（请读者想一想这是为什么），因此有 $\Delta_a(\varphi)(\partial V/\partial \varphi_a) = 0$ （对相同指标求和，下同）。将这一表达式对 φ_b 作一次导数，并注意到真空所满足的条件，可得（请读者自行证明）：

$$\Delta_a(\varphi) \frac{\partial^2 V}{\partial \varphi_a \partial \varphi_b} = 0$$

由上式可以看到，每一个 $\Delta_a(\varphi) \neq 0$ 的连续对称变换都对应于的一个本征值为零的本征态，从而也就对应于一个无质量标量粒子。而另一方面， $\Delta_a(\varphi) \neq 0$ 的连续对称变换所对应的正是那些不能使真空态不变——从而被真空态所破缺（即自发破缺）——的连续对称性。这就证明了每一个自发破缺的整体连续对称性都必然伴随一个无质量标量粒子，即戈德斯通粒子。这正是戈德斯通定理。^[16]（请读者思考一下，戈德斯通定理中的“整体”二字体现在证明的什么地方？）由于自发破缺的整体连续对称性的数目等于这些对称性的生成元的数目，因此戈德斯通定理也表明了戈德斯通粒子的数目等于自发破缺的整体连续对称性生成元的数目。举个例子来说，SU(2)对称性具有三个生成元，若完全破缺，就会产生三个戈德斯通粒子；若破缺为U(1)，则只产生两个戈德斯通粒子（因为有一个生成元未破缺）。进一步的分析还表明，戈德斯通粒子与那些自发破缺的整体连续对称性所对应的荷——关于荷，请

读者回忆一下诺特定理（Noether theorem）——具有相同的宇称及内禀量子数。

当然，严格讲，上面的证明只是在所谓经典层次上的证明，而没有考虑量子修正。那么考虑了量子修正后，戈德斯通定理是否仍成立呢？答案是肯定的，而且证明也基本一样，只需用包含量子修正的所谓量子有效势 V_{eff} 取代经典拉氏量中的势函数 V 即可^[17]。

由戈德斯通等人证明的这一结果为什么会对把近似对称性归因于对称性自发破缺的想法造成致命打击呢？原因很简单，那就是近似对称性中的某一些——比如同位旋对称性——正是整体连续对称性，如果它们的近似性果真源自对称性自发破缺，那就应该存在相应的无质量标量粒子。但我们从未在实验上观测到任何这样的粒子。因此对称性自发破缺的想法在粒子物理学中由于牵涉到无质量粒子而陷入了困境。

八、从希格斯机制到电弱统一理论

无独有偶，粒子物理学中产生于20世纪五六十年代的另一个很高明的想法也受到了无质量粒子的困扰，那个想法是1954年由杨振宁（1922—）和米尔斯（Robert Mills, 1927—1999年）提出的，现在被称为杨-米尔斯理论（Yang-Mills theory）。这是一种所谓的定域“非阿贝尔规范理论”（non-Abelian gauge theory），是对像量子电动力学那样的定域“阿贝尔规范理论”（Abelian gauge theory）的推广^[18]，具体的区别是以非阿贝尔规范对称性取代了量子电动力学所具有的阿贝尔规范对称性——即 $U(1)$ 规范对称性。提出这种理论最初的动机是企图用它来描述同位旋对称性。但这一企图立刻就遇到了一个很大的困难，那便是这种理论所具有的定域规范对称性会无可避免地导致无质量的矢量粒子（被称为规范粒子，类似于量子电动力学中的光子），而在现实中，除光子外我们从未在实验上观测到任何这样的粒子。

就这样，杨-米尔斯理论与对称性自发破缺这两个出色的想法先后搁浅了，推根溯源，都是无质量粒子惹的祸。但如果我们仔细研究一下这对“难兄难弟”的病根，就会发现两者竟然像是互为解药！对称性自发破缺的问题出在哪里呢？出在整体连续对称性上；而杨-米尔斯理论的问题又出在哪里呢？出在定域规范对称性（那是一种特殊的定域连续对称性）上。如果我们把这两者放在一起，让对称性自发破缺干掉那些产生无质量矢量粒子的定域规范对称性，杨-米尔斯理论不就可以摆脱困境了吗？更妙的是，由于杨-米尔斯理论中的对称性不是整体而是定域的，戈德斯通定理将不适用于这种对称性的自发破缺，这样一来说不定那些可恶的戈德斯通粒子也会消失，那岂不是两全其美？世界上会有这么好的事吗？还真的有。

最早明确指出这一点的是美国凝聚态物理学家安德森（Philip Warren Anderson, 1923—）。对于安德森来说，戈德斯通定理显然不可能是普遍成立的，因为当时的凝聚态物理学家们已经知道，超导体就是一个连续对称性—— $U(1)$ 对称性——自发破缺的体系，但在这一破缺的过程中并没有产生无质量的戈德斯通粒子。安德森并且很正确地意识到了 $U(1)$ 对称性的定域特点是使戈德斯通定理失效的关键。由于并非只有定域 $U(1)$ 对称性具有定域特点，事实上所有杨-米尔斯理论也都具有这一特点。因此安德森在1963年猜测道：“戈德斯通的零质量困难并不是一个严重的困难，因为我们很可能可以用一个相应的杨-米尔斯零质量问题来消去它。”安德森的想法得到了一些物理学家的认同，但也有人认为这种凝聚态物理的类比不能应用到相对论量子场论中。

这种怀疑很快就被推翻了。1964年，英国物理学家希格斯（Peter Higgs, 1929—）、比利时物理学家英格勒特（François Englert, 1932—）与布罗特（Robert Brout, 1928—）等几乎同时证实了安德森的想法。这便是描述规范对称性自发破缺的著名的希格斯机制（Higgs mechanism），它一方面消除了无质量的戈德斯通粒子，另一方面则使规范粒子获得了质量^[19]。

不过希格斯等人的漂亮工作并没有引起即刻的轰动。希格斯就这一工作所写的两篇短文中的第二篇甚至一度遭到了退稿，理由是“与物理世界没有明显关系”。这一退稿理由使希格斯深感不快，但也促使他更深入地考虑了理论可能引致的实验结果，并对论文进行了补充。希格斯后来认为，他因遭到退稿而补充的那些内容是人们将希格斯粒子及希格斯机制与他的名字联系在一起的主要原因。

做了这么多背景介绍，现在让我们回到主题——质量的起源——上

来。希格斯机制不仅一举“救活”了粒子物理学中对称性自发破缺与杨-米尔斯理论这两个极为出色的想法，而且在救助过程中为我们提供了一种产生质量的新方法，即通过规范对称性的自发破缺，从不带质量项的拉氏量中产生出质量来。不过，由此而获得质量的——如上文及注释所述——只是规范粒子，而规范粒子的质量在宇宙可见物质的质量中只占了微不足道的比例，我们更关心的是在可见物质质量中占主要比例的那些粒子——费米子。

那么，费米子的情况如何呢？1967年，温伯格和萨拉姆将希格斯机制应用到美国物理学家格拉肖（Sheldon Lee Glashow, 1932—）等人几年前所提出的旨在描述电磁和弱相互作用的SU（2）×U（1）规范理论中，建立起了所谓的电弱统一理论（electroweak theory）^[20]。这一理论与描述强相互作用的量子色动力学（quantum chromodynamics）一起组成了粒子物理的标准模型。在标准模型中，费米子也是通过规范对称性的自发破缺——或者更确切地说，通过电弱统一理论中的规范对称性自发破缺——获得质量的。具体地讲，在标准模型中，费米场 ψ 与希格斯机制中的标量场（也称为希格斯场） ϕ 之间存在所谓的汤川耦合（Yukawa coupling）： $-\lambda \bar{\psi} \phi \psi$ （其中 λ 为耦合常数^[21]）。由于希格斯场具有非零的真空期待值，因此将这一耦合项相对于真空展开后就会出现形如 $-m \bar{\psi} \psi$ 的费米子质量项。

因此，我们可以说，标准模型中所有基本粒子的质量都来源于电弱统一理论中的规范对称性自发破缺。这是标准模型对质量起源问题的直接回答。

不过遗憾的是，这一回答却是一个不尽人意的回答。为什么这么说呢？因为这一回答从某种意义上讲与其说是回答了问题，不如说是在转

嫁问题——把我们想要理解的基本粒子的质量转嫁给了希格斯场的真空期待值、规范耦合常数以及汤川耦合常数。这其中希格斯场的真空期待值及规范耦合常数与基本粒子——主要是费米子——的种类无关，可以算是具有普适性的，因此将质量向这些参数约化不失为是一种有效的概念约化。但汤川耦合常数则不然，它对于每一种费米子都有一个独立的数值。由于这些参数的存在，标准模型的拉氏量虽然不显含质量参数，但它所包含的与质量直接有关的自由参数的数目却一点也不比原先需要解释的质量参数的数目来得少（事实上还略多一点）。从某种意义上讲，用这种方式来解释质量的起源，就像英国物理学家霍金（Stephen Hawking, 1942—）在《时间简史》（*A Brief History of Time*）一书中引述的一位老妇人的“理论”。那位老妇人宣称世界是平面的，由一只大乌龟托着。当被问到那只大乌龟本身站在哪里时，老妇人冷静地回答说：“站在另一只大乌龟的背上。”

因此，希格斯机制及包含希格斯机制的电弱统一理论虽然从许多唯象的方面来衡量是非常成功的，其所体现的把质量与真空的对称性破缺性质联系在一起的思路也极为深刻。但它们作为与对称性破缺有关的特殊机制或模型，本身却没能实现对质量概念的真正约化，从而不能被认为是质量起源问题令人满意的回答。

九、量子色动力学

与戈德斯通、希格斯等人在对称性自发破缺方面的研究几乎同时，物理学家们在研究强相互作用上也取得了重大进展。1961年，美国物理学家盖尔曼（Murray Gell-Mann, 1929—）与以色列物理学家内曼（Yuval Ne'eman, 1925—2006年）彼此独立地提出了强子分类的SU（3）模型^[22]。这一模型不仅对当时已知的强子给出了很好的分类，而且还预言了当时尚未发现的粒子，比如 Ω -粒子^[23]。但这一模型有一个显著的缺陷，那就是SU（3）的基础表示（fundamental representation）似乎不对应于任何已知的粒子。1964年，盖尔曼与美国物理学家茨威格（George Zweig, 1937—）提出了夸克（quark）模型，将夸克作为SU（3）基础表示所对应的粒子，强子则被视为是由夸克组成的复合粒子^[24]。

在夸克模型中，为了给出正确的强子性质，夸克必须具有实验上从未发现过的量子数，比如分数电荷，这在当时是令人不安的。对此，盖尔曼也深感困惑，只能用“夸克存在但不是真实的”（they exist but are not real）这样诡异的语言来搪塞。夸克模型的另一个麻烦是，夸克是费米子，而某些强子却似乎包含三个处于同一量子态的夸克，从而违反了泡利不相容原理（Pauli exclusion principle）。关于这一点，1965年美国物理学家格林伯格（Oscar W. Greenberg, 1932—）、韩国物理学家韩武永（Moo-Young Han, 1934—）和南部阳一郎先后提出了一个解决方案，那就是引进一个新的三值量子数以保证那些夸克具有不同的量子态。南部阳一郎甚至粗略地设想了以这一量子数为基础构造杨-米尔斯理论，但这些工作并未引起重视。1972年，盖尔曼等人在实验的引导下重新考虑了这一被盖尔曼称之为色荷（color）的新量子数，以及以之为

基础的杨-米尔斯理论。这一理论被称为了量子色动力学。由于色荷是一个三值量子数，因此量子色动力学的规范群被选为了SU（3）。

在量子色动力学的发展过程中，20世纪60年代末的一系列所谓“电子-核子深度非弹性散射”（deep-inelastic electron-nucleon scattering）实验起了很大的作用。这些实验不仅证实了核子内部存在着点状结构，而且还显示出这些点状结构之间的相互作用在高能——即近距离——下会变弱。这些点状结构被美国物理学家费恩曼（Richard Feynman, 1918—1988年）称为“部分子”（parton），它们中的一部分后来被证实就是夸克（另一部分是后面会提到的胶子），而部分子之间的相互作用在高能——即近距离——下变弱的行为则被称为渐近自由（asymptotic freedom）。渐近自由为实验上从未观测到孤立夸克这一事实提供了一种很好的说明：那就是当夸克彼此远离时，它们之间的相互作用会越来越强，最终从真空中产生出足以中和它们所带色荷的粒子。我们在实验上能够分离出的任何粒子——比如强子——都只能是这种色荷中和之后的产物，而不可能是孤立的夸克^[25]。由于这一原因，渐近自由很快被视为描述夸克相互作用的理论所必须具备的性质。

1973年，美国物理学家波利策（Hugh David Politzer, 1949—）、韦尔切克（Frank Wilczek, 1951—）和格娄斯（David Gross, 1941—）等人发现杨-米尔斯理论具有渐近自由性质^[26]。在当时已知的所有四维可重整场论中，杨-米尔斯理论是唯一具备这一性质的理论，这对盖尔曼等人提出的量子色动力学是一个很强的支持。那时候，人们对杨-米尔斯理论本身的研究也已取得了系统性的进展：1967年，苏联物理学家法捷耶夫（Ludvig Faddeev, 1934—）和波波夫（Victor Popov, 1937—1994年）完成了杨-米尔斯理论的量子化；1971年，荷兰物理学家特·胡夫特（Gerard't Hooft, 1946—）证明了杨-米尔斯理论的可重整性。在这一系

列工作的基础上，量子色动力学顺理成章地成为了标准模型中描述强相互作用的基本理论。这一理论中对应于SU（3）生成元的八个载力子被称为胶子（gluon），它们都是无质量的。

看到这里，有些读者可能会问：我们是不是离题了？量子色动力学中总共只有两类粒子：胶子与夸克。其中胶子是无质量的，而夸克虽然有质量，但其质量——与标准模型中其他费米子的质量一样——却是由电弱统一理论中的规范对称性自发破缺产生的，与量子色动力学无关。既然如此，量子色动力学与质量起源这一主题又能有什么关系呢？应该说，这是一个很合理的疑问。但量子色动力学的奇妙之处就在于，它形式上异常简洁——一个简简单单的规范群，一个平平常常的耦合常数，差不多就是全部的家当——但内涵却惊人地丰富。它宛如一坛绝世的佳酿，越品就越是回味无穷。在谈论质量起源问题的时候，人们往往把注意力放在希格斯机制及包含希格斯机制的电弱统一理论上——因为希格斯机制在登场伊始就打出了质量产生机制的响亮广告。但事实上我们将会看到，看似与质量起源问题无关的量子色动力学对这一问题有着非常独特而精彩的回答，而且从某种意义上讲，这一回答才是标准模型范围内的最佳回答。

我们先来看看量子色动力学的拉氏量：

$$L = -\frac{1}{2}\text{Tr}(G^{\mu\nu}G_{\mu\nu}) + \sum_q \bar{q}(i\gamma^\mu D_\mu - m_q)q$$

其中 q 为夸克场； $G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu]$ 为规范场强； D_μ 为协变导数； A_μ 为规范势； m_q 为夸克 q 的质量； g 为耦合常数；式中的求和遍及所有的夸克种类。自然界已知的夸克种类——也称为“味”（flavor）——共有六种。其中u（上夸克）、d（下夸克）、s（奇夸克）被称为

轻夸克，质量分别约为 2.3MeV 、 4.8MeV 和 95MeV ； c （粲夸克）、 b （底夸克）、 t （顶夸克）被称为重夸克，质量分别约为 1.3GeV 、 4.2GeV 和 173GeV 。这其中轻夸克的质量是在约 2GeV 的能标上定义的，重夸克的质量则是在其自身质量标度上定义的^[27]。这些质量参数本身在标准模型范围内是不能约化的，但由这些夸克所组成的强子的性质，在很大程度上可以由量子色动力学来描述，这其中就包括强子的质量。

在接下来的几节中，我们就来看一下量子色动力学对强子质量的描述，以及这种描述在何种意义上可以被视为是对质量起源问题的回答。

十、同位旋与手征对称性

我们知道，可见物质的质量主要来自于质子和中子，其中质子由两个u夸克及一个d夸克组成，而中子由一个u夸克及两个d夸克组成。在下面的叙述中，我们将只考虑这两种夸克。由于这两种夸克的质量远小于包括质子和中子在内的任何强子的质量，作为近似，我们先忽略它们的质量。这时量子色动力学的拉氏量为

$$L = -\frac{1}{2}\text{Tr}(G^\mu G_\mu) + i\bar{u}\gamma^\mu D_\mu u + i\bar{d}\gamma^\mu D_\mu d$$

显然（请读者自行验证），这一拉氏量在以下两个整体SU（2）变换：

$$\psi \rightarrow \exp(-it^a\theta^a)\psi, \quad \psi \rightarrow \exp(-i\gamma^5 t^a\theta^a)\psi$$

下是不变的。这其中 $\psi = (u, d)^T$ ， t^a 是SU（2）的生成元（即泡利矩阵的1/2）。这两个存在于u夸克和d夸克之间的对称性分别被称为同位旋对称性与手征对称性（chiral symmetry），记为SU（2）_V与SU（2）_A。这其中同位旋对称性SU（2）_V只要夸克质量彼此相等（不一定要为零）就存在，而手征对称性SU（2）_A只有在夸克质量全都为零时才具有（这一情形因此而被称作手征极限）。这一点与我们在第六节中提到的无质量量子电动力学的手征对称性类似。除此之外，这一拉氏量还存在一个显而易见的整体U（1）_V对称性，它对应于重子数守恒，与夸克是否有质量，以及质量是否彼此相等都无关。

综合起来，忽略夸克质量的上述拉氏量具有整体SU（2）_V×SU（2）_A×U（1）_V对称性^[28]。在这些对称性中，同位旋对

称性SU(2)_V与手征对称性SU(2)_A所对应的守恒流分别为

$$V^{\mu a} = \bar{\psi} \gamma^\mu t^a \psi, \quad A^{\mu a} = \bar{\psi} \gamma^\mu \gamma^5 t^a \psi$$

显然，在宇称变换下， $V^{\mu a}$ 是矢量（vector）， $A^{\mu a}$ 则是轴矢量（axial vector）。它们对应的荷 $(Q_V)^a = \int V^{0a} d^3x$ 与 $(Q_A)^a = \int A^{0a} d^3x$ 分别为标量（scalar）及赝标量（pseudoscalar）^[29]。

如果同位旋与手征对称性都是严格的对称性，那么 $(Q_V)^a$ 将生成强子谱中自20世纪60年代起逐步引导人们发现量子色动力学的同位旋对称性；而 $(Q_A)^a$ 则将生成所谓的手征对称性，它要求每一个强子都伴随有自旋、重子数及质量与之相同，而宇称却相反的粒子——那样的对称性在强子谱中并未被发现过。

对此，最容易想到的解释是：由于u夸克和d夸克实际上并不是无质量的，因此手征对称性本就不可能严格成立。事实上，不仅手征对称性不可能严格成立，由于u夸克和d夸克的质量彼此不同，连同位旋对称性也不可能严格成立。但是，考虑到u夸克和d夸克的质量相对于强子质量是如此之小，相应的对称性在强子谱中似乎起码应该近似地存在。对于同位旋对称性来说，情况的确如此（否则就不会有早年那些强子分类模型了）^[30]。但手征对称性却哪怕在近似意义上也根本不存在。举个例子来说，手征对称性要求介子三重态 $\rho(770)$ 与 $a_1(1260)$ 互为对称伙伴（请读者自行查验这两组介子的量子数），但实际上这两者的质量分别约为775MeV和1230MeV^[31]，相差悬殊（作为对比，同位旋伙伴的质量差通常都在几个MeV以下），连近似的对称性也不存在。

初看起来，事情似乎出了麻烦，但物理学家们却从这一麻烦中找到了一条探究低能量量子色动力学的捷径。正所谓“山重水复疑无路，柳暗花明又一村”。

十一、手征对称性自发破缺

手征对称性 $SU(2)_A$ 是量子色动力学拉氏量中的（近似）对称性，却在现实世界中完全找不到对应，这究竟是什么原因呢？应该说，要猜测一下是不困难的，因为当时物理学家们已经知道对称性可以自发破缺。如果量子色动力学中的手征对称性是自发破缺的，显然就会出现这种拉氏量具有（近似）手征对称性，现实世界却不并不买账的现象。但是，猜测归猜测，要想在理论上严格证明这一点——哪怕只是在物理学而不是数学的标准下严格证明——却是极其困难的。

有读者可能会问：对称性自发破缺在电弱统一理论中用得好好的，为什么在量子色动力学中却变得“极其困难”了呢？这是因为在电弱统一理论中对称性自发破缺是由人为引进的希格斯场产生的，我们有一定的自由度来选择对称性破缺的方式。但量子色动力学并不包含这种人为引进的希格斯场，因此，在量子色动力学中，整体 $SU(2)_V \times SU(2)_A \times U(1)_V$ 对称性是否自发破缺？如果破缺，是否恰好是手征部分 $SU(2)_A$ 破缺，即破缺到 $SU(2)_V \times U(1)_V$ ？都只能由理论本身来决定，而不是我们可以擅自假设的，正是这一特点使问题变得“极其困难”^[32]。更麻烦的是，手征对称性的破缺——如果出现的话——乃是一种出现在量子色动力学的强相互作用区域——即低能区域——的现象。对于理论研究来说，这无疑是雪上加霜。

另一方面，对称性自发破缺的存在与否及具体方式由理论本身所决定，虽然为量子色动力学带来了一个“极其困难”的理论问题，同时却也是它的一个极大的理论优势。因为电弱统一理论之所以只是对质量起源问题的一个不尽人意的回答，一个很重要的原因就是希格斯场以

及它与费米场之间的相互作用——汤川耦合——都是人为引进的，从而都是所谓的自由参数（free parameter）。而量子色动力学没有那种类型的自由参数，因此它与观测之间的对比更为严酷：如果成功，将是极具预言能力的成功，因为自由参数越少，预言能力就越强；但如果失败，也将是无力回天的失败，因为自由参数越少，回旋余地也就越小。

那么量子色动力学究竟能不能实现从 $SU(2)_V \times SU(2)_A \times U(1)_V$ 到 $SU(2)_V \times U(1)_V$ 的对称性自发破缺呢？目前在理论上还是一个待解之谜。1979年，特·胡夫特通过对规范理论中的反常（anomaly）进行分析，得到了一个结果：即如果所考虑的整体对称性是 $SU(3)_V \times SU(3)_A \times U(1)_V$ ，那它就必须自发破缺。可惜的是，一来量子色动力学中的 $SU(3)$ 对称性远比 $SU(2)$ 对称性粗糙，二来这一结果并未告诉我们具体哪一部分对称性会自发破缺。1980年，美国物理学家科尔曼（Sidney Coleman, 1937—2007年）与威顿（Edward Witten, 1951—）提出了在某些合理的物理条件下，当色的数目 N_c 趋于无穷时，手征对称性必须自发破缺。这一结果虽然抓准了手征对称性，但可惜量子色动力学中色的数目 N_c 不仅不是无穷，而且还很小（ $N_c=3$ ）。1984年，伊朗裔美国物理学家瓦法（Cumrun Vafa, 1960—）与威顿证明了未被非零夸克质量项所破缺的同位旋对称性（请读者想一想，在现实世界里这一对称性由什么群来表示？）不会自发破缺。可惜这一证明虽然表明特定的同位旋对称性不会自发破缺，却未能对手征对称性是否一定会自发破缺提供说明。

虽然上述理论研究没有一个能够证明量子色动力学中的 $SU(2)_V \times SU(2)_A \times U(1)_V$ 整体对称性必定会自发破缺到 $SU(2)_V \times U(1)_V$ ，但它们都与这一破缺方式相容这一事实，无疑还

是大大增强了人们的信心。在物理学上，严格证明是一种美妙的东西，但有时却可望不可及，物理学家们的工作往往并不总是依赖于它。迄今为止，虽然尚未有人能够给出量子色动力学中手征对称性自发破缺的严格证明，但从这一破缺方式已经得到的大量间接证据来看，它的证明应该只是时间问题。物理学家们更感兴趣的是：如果手征对称性自发破缺，我们可以从中得到什么推论？有关这一点，人们做过不少细致研究。那些研究获得了极大的成功，不仅给出了被称为“手征微扰理论”（chiral perturbation theory）的描述低能量量子色动力学的所谓“有效场论”（effective field theory），而且得到了一系列与实验相吻合的漂亮结果。这一切也反过来为手征对称性的自发破缺提供了进一步的间接证据。

下面我们就来看看由手征对称性自发破缺导致的推论之中与质量起源问题有密切关系的部分。

十二、赝戈德斯通粒子的质量

我们在第七节中介绍过，对称性自发破缺的最重要的推论之一，是存在无质量的标量粒子，即戈德斯通粒子，它们与自发破缺的对称性所对应的荷具有相同的宇称及内禀量子数。对于手征对称性来说，荷是 $(Q_A)^a$ ，它在时空中是一组赝标量，在内禀空间中则是一个矢量，因此相应的戈德斯通粒子的宇称为负，同位旋则为1。自然界里满足这些特征的强子中质量最轻的是 π 介子（ π^- 、 π^0 和 π^+ ）。如果手征对称性是自发破缺的， π 介子就应该是这一破缺所对应的戈德斯通粒子^[33]。但是，戈德斯通粒子是无质量的， π 介子却是有质量的，这一矛盾该如何解决呢？

我们知道，在理想的对称性自发破缺情形下，体系的实际真空态可以是一系列简并真空态中的任何一个。但是，量子色动力学中的手征对称性破缺却并非理想情形下的破缺，因为量子色动力学的拉氏量含有手征对称性的明显破缺项——即夸克的质量项。由于这种明显破缺项的存在，实际真空态的选取就不再是任意的了，明显破缺项的存在将会对实际真空态起到一个选择作用。这就好比一根立在桌上的筷子，如果桌子是严格水平的，它向任何一个方向倒下都是同等可能的，但如果桌子是倾斜的，它就会往倾斜度最大（梯度最大）的方向倒。用数学的语言来说（符号的含义与第七节相同），如果 $V_1(\phi_a)$ ($a=1,2,\dots,N$) 表示对称性的明显破缺项，那么，它所选出的真空态将满足下列条件：

$$\Delta_a(\phi) \frac{\partial V_1}{\partial \phi_a} = 0$$

这一条件被称为真空取向条件（vacuum alignment condition）。另

一方面，明显破缺项的存在也破坏了戈德斯通定理成立的条件，由此导致的结果是戈德斯通粒子有可能具有非零质量，这样的粒子被称为赝戈德斯通粒子（pseudo-Goldstone particle）。真空取向条件是确定赝戈德斯通粒子质量的重要条件。赝戈德斯通粒子的出现消除了 π 介子的非零质量与戈德斯通粒子的零质量之间的定性矛盾。但在定量上 π 介子与赝戈德斯通粒子的质量是否吻合呢？我们现在就来看一看。

如前所述，对于量子色动力学中的手征对称性来说，对称性的明显破缺项为质量项，它可以改写成（请读者自行验证）：

$$V_1 = \frac{1}{2}(m_u + m_d)\bar{\phi}\phi + \frac{1}{2}(m_u - m_d)(\bar{u}u - \bar{d}d)$$

其中 $\bar{\phi}\phi = \bar{u}u + \bar{d}d$ 。上式的特点是：第一项只破坏手征对称性，第二项则破坏同位旋对称性。研究表明，在这些特点的基础上进一步考虑到不存在同位旋对称性的自发破缺这一限制，可以得到赝戈德斯通粒子的质量为（这一结果也可以从手征微扰理论得到）：

$$M_\pi^2 = \frac{m_u + m_d}{2F_\pi^2} \langle 0 | \bar{\phi}\phi | 0 \rangle$$

其中 F_π 是一个量纲为能量的常数，由

$$\langle 0 | A^{\mu a}(x) | \pi^b(p) \rangle = i p^\mu F_\pi \delta^{ab} e^{-ipx}$$

定义。 F_π 被称为 π 衰变常数（pion decay constant），可以由 π 介子的衰变来确定，原则上也可以从理论上计算出，其数值约为92.4MeV^[34]。是一个量纲为能量三次方的参数，被称为手征凝聚（chiral condensation），目前人们对它的计算还比较粗略，结果大致为，其中 n_f 为参与凝聚的夸

克种类，对于我们所考虑的情形 $n_f=2$ （即只有u夸克和d夸克参与凝聚）^[35]。 m_u+m_d 通常取为 $8\sim 9\text{MeV}$ 。由此可以得到（请读者自己计算一下）： $m_\pi\sim 140\text{MeV}$ 。这几乎正好就是 π 介子的质量（ π^\pm 的质量约为 140MeV ； π^0 的质量约为 135MeV ）。当然，上述估算是相当粗略的，不能因为数值上的吻合而高估它的精度。但结合了格点量子色动力学（lattice QCD）计算的大量更为细致的研究表明，这种吻合并非偶然^[36]。

现在让我们再次回到主题——质量的起源——上来。我们看到，量子色动力学计算出了作为赝戈德斯通粒子的 π 介子的质量。如果我们想知道 π 介子的质量起源，这可以算是一种回答。可惜的是，这种回答与我们在第六节中介绍的电磁自能具有相同的缺陷，那就是它正比于在理论中无法约化的外来参数：夸克质量。一旦外来参数不存在（即夸克质量为零），这一回答就会失效（因为答案也将为零）。因此量子色动力学对 π 介子及其他赝戈德斯通粒子质量的计算虽然很漂亮，从回答本原问题的角度看却仍不足以令人满意。

十三、一个**93**分的答案

但是，当我们把目光转到更复杂，同时也更具现实意义的强子——比如质子和中子（以下合称核子）——的质量时，却会看到量子色动力学的确为质量起源问题提供了一个非常精彩的回答。

计算核子或其他重子的质量是一个相当困难的低能量量子色动力学问题，通常的做法是利用巨型计算机进行格点量子色动力学计算。但是，由于技术上的限制，人们在这类格点量子色动力学计算中采用的u夸克和d夸克的质量一度要比它们的实际质量高出5倍左右，由此得到的核子质量通常也要比实际值高出30%以上。不过近几年，随着技术的演进，格点量子色动力学计算所采用夸克质量已逐渐降低，甚至已有一些研究者开始采用实际质量。

另一方面，与格点量子色动力学计算中夸克质量的“不可承受之重”截然相反，在我们前面提到的手征微扰理论中，夸克的质量却是越轻越好，甚至最好是零。显然，如果我们能在这两种极端之间作某种调和，借助手征微扰理论对格点量子色动力学的计算进行适当的外推，就有可能得到更接近现实世界的结果。这正是物理学家们在计算核子质量时采用的手段。这种借助手征微扰理论对格点量子色动力学计算进行外推的方法被称为手征外推（chiral extrapolation）。利用手征外推得到的核子质量为

$$m_N = m_0 - 4c_1 m_\pi^2 + O(m_\pi^3)$$

其中 $m_0 \approx 880 \text{ MeV}$; $c_1 \approx -1 \text{ GeV}^{-1}$; m_π^2 是 π 介子的质量平方，如上节所述，正比于夸克质量。若干更高阶的项也已被计算出，这里就不细述了。将

有关数据代入这一公式，我们可以得到（请读者自己计算一下）： $m_N \approx 954 \text{ MeV}$ ，它与实际的核子质量（质子约为 938 MeV ；中子约为 940 MeV ）相当接近。不仅如此，系统的计算（包括来自部分高阶项的贡献）还给出了许多其他重子的质量，比如： $m_\Sigma \approx 1192 \text{ MeV}$ （实验值约为 $\Sigma^+ : 1189 \text{ MeV}$ ； $\Sigma^0 : 1193 \text{ MeV}$ ； $\Sigma^- : 1197 \text{ MeV}$ ）； $m_\Lambda \approx 1113 \text{ MeV}$ （实验值约为 1116 MeV ）； $m_\Xi \approx 1319 \text{ MeV}$ （实验值约为 $\Xi^0 : 1315 \text{ MeV}$ ； $\Xi^- : 1321 \text{ MeV}$ ），都与实验有不错的吻合^[37]。这些结果表明，量子色动力学的确可以用来计算重子质量。

那么，从回答本原问题的角度看，这些计算是否令人满意呢？

从上面所引的核子质量公式中我们可以看到，上述核子质量有一个不同于赝戈德斯通粒子质量的至关重要的特点，那就是它在手征极限——即夸克质量为零——时不为零，而等于 $m_0 \approx 880 \text{ MeV}$ 。这个数值约为核子质量的93%，它完全是由量子色动力学所描述的相互作用所确定的^[38]。这表明，即便不引进任何外来的夸克质量，量子色动力学仍能给出核子质量的绝大部分。由于宇宙中可见物质的质量主要来自核子质量，因此宇宙中可见物质质量的绝大部分都可以在不引进夸克质量的情况下，由纯粹的量子色动力学加以说明。从这个意义上讲，量子色动力学为质量起源问题提供了一个独特而精彩的回答。这一回答不像电弱统一理论那样带有比所要解释的质量参数还要多的可调参数，因而非常符合回答本原问题的需要。不过，由于它只能给出核子质量的93%，因此我们粗略地给它打93分。在标准模型的范围内，这是迄今所知的最佳回答。

93分虽然是一个高分，但终究不是满分。为了寻找更接近满分的答

案，我们不得不重新回到标准模型中不能约化的那些质量——包括使量子色动力学丢掉7分的夸克质量——上来。那些质量究竟来自何方？究竟还能不能约化？这些问题的答案——如果有的话——就只能到标准模型之外去寻找了。

2007年1月25日写于纽约

2014年11月19日最新修订

[1]本系列曾在《现代物理知识》杂志（中国科学院高能物理研究所）上连载，其中第3~6节发表于2007年第1期（发表时的标题为：《质量起源——电磁质量说的兴衰》）；第7~8节发表于2007年第2期（发表时的标题为：《质量起源——从对称性破缺到希格斯机制》）；第9~13节发表于2007年第3期（发表时的标题为：《质量起源——量子色动力学与质量起源》）。

[2]当然，这一说法并不严格，在星系所占据的空间范围内也有数量可观的暗物质及暗能量，我们这里指的只是光学观测意义上的星系。

[3]这里有一个著名的例外是马赫（Ernst Mach, 1838—1916年），他对牛顿绝对时空观的批判性思考启示了这样一种观念，那就是一个物体的质量（惯性）起源于宇宙中其他星体的作用。马赫的想法曾对爱因斯坦产生过影响，并且直到现在还有一些物理学家在研究，但它与广义相对论的定量结果及对惯性各向异性的测量结果并不相符。因此我们不把它列为有关质量起源的具体理论。

[4]洛伦兹所用的质量定义是 $m \left(\frac{dv}{dt} \right) = dp/dt$ ，“横质量”与“纵质量”分别对应于 v 与 dv/dt 垂直及平行这两种特殊情况。

[5]当时还没有爱因斯坦的质能关系式，亚伯拉罕的这一关系式是一个简单的力学关系式，读者不妨自行推导一下。

[6]如上所述，亚伯拉罕也是经典电子论的代表人物，有读者可能会问，他自己的电子模型又如何呢？与洛伦兹不同，亚伯拉罕所用的是一个绝对刚性的电子模型，因此在他的模型中不需要引进对能量有贡献的张力。他的模型一度曾被认为比洛伦兹的模型更符合实验，但那实验——即德国物理学家考夫曼（Walter Kaufmann, 1871—1947年）的实验——后来被证实是有缺陷的。

[7]有兴趣的读者可以进一步证明这样一些结果：（1）对于球对称均匀面电荷分布， $\int T_{\text{EM}}^{\text{ij}}(x^{\text{e}}) d^3x = (1/3) \int T_{\text{EM}}^{00}(x^{\text{e}}) d^3x$ ；（2）对于任意球对称电荷分布， $\int T_{\text{EM}}^{\text{ij}}(x^{\text{e}}) d^3x = (1/3) \int T_{\text{EM}}^{00}(x^{\text{e}}) d^3x$ ；（3）由1和2证明洛伦兹有关 m_{T} 与 m_{L} 的公式；（4）证实亚伯拉罕对洛伦兹的批评，即用 $m_{\text{L}} = (1/v) (dE/dv)$ 定义的质量与洛伦兹的结果不同。

[8]当然，这样的说法对历史作了一定的简化。确切地讲，经典电子论的出现实际上略早于电子的发现，而类似于经典电子论的电子结构研究在量子理论之后仍间或地有一些物理学家在做，不过那些研究大都已不能完全归于经典电子论的范畴。另一方面，经典电子论所包含的电子结构以外的东西，比如从物质的微观——但非量子的——电磁结构出发研究宏观电磁及光学性质的方法，直到今天仍可以在一些经典电磁学的教材中找到踪迹。但总体来说，经典电子论随着量子理论的兴盛而没落的大趋势仍是显而易见的。

[9]经典电子论对电子的描述不仅与量子力学不符，在电子自旋发现之后，试图在经典电子模型中加入电子自旋的努力与狭义相对论也产生了矛盾，可谓腹背受敌。

[10]韦斯科夫的计算包含了一个符号错误，但很快被弗里（Wendell H. Furry, 1907—1984年）和卡尔森（Frank Carlson）所纠正。

[11]量子场论的微扰展开式有许多微妙的地方。以量子电动力学为例，尽管其耦合常数 α 很小，从而 n 圈图的贡献受到 α^n 的抑制，但另一方面，随着圈数的增加，不等价 n 圈图的数目也在增加，其趋势约为 $n!$ （这当然只是非常粗略的说法，圈图的确切数目与相互作用的具体形式有关，且其中还有符号问题，综合的结果非常复杂）。当 n 接近或大于 $1/\alpha$ 时，圈图数量的增加将抵消由弱耦合所带来的减弱因子 α^n 的影响，因此量子电动力学的微扰展开式并不收敛，这一点

最早是由英裔美国物理学家戴森（Freeman Dyson）于1951年给出的。有鉴于此，所谓单圈图的贡献占了主要部分其实是从渐近级数的意义上说的。

[12]顺便提一下，庞加莱张力带来的困难除了我们在第四节中提到的非电磁起源外，还有一个更严重的，那就是由庞加莱张力所维持的电子结构虽然具有静态的平衡，却是不稳定的，在细微的扰动下就会土崩瓦解（类似于爱因斯坦的静态宇宙模型）。这是1922年由意大利物理学家费米（Enrico Fermi）所证明的。

[13]当我们谈到截断的时候，有一点需要提醒读者注意，那就是对于像电子自能这样对截断能标相对敏感的物理量，只计算截断能标以下的贡献显然是不完整的，那么来自截断能标以上的贡献有多少呢？答案是与适用于截断能标以上的理论的具体形式有关。如果那个理论本身也有截断，我们还必须关心来自那个截断能标以上的贡献。物理学家们的期望是，我们最终将会会有一个有限的理论，那时我们就不需要用截断来遮遮掩掩了。

[14]这从简单的量纲分析就可以看出： δm 的形式为 $mf(\Lambda/m)$ ，而从费恩曼图所对应的积分的形式可知其相对于 Λ 的渐近形式 $f(x)$ 只能是对数或以正负整数为幂次的幂函数，这其中只有 $f(x) = \ln(x)$ 可以使 δm 既在 $\Lambda \rightarrow \infty$ 时发散，又在 $m \rightarrow 0$ 时为零。

[15]学过量子力学的读者可能会进一步问：如果一个量子体系的基态是简并的，那么体系的物理基态难道不应该是这些简并态的某种量子叠加吗？这种量子叠加——如我们在量子力学中所见到的——往往不仅会破除原有的基态简并性，并且使真正的基态具有与原先简并基态的集合相同的对称性。在这种情况下，对称性自发破缺岂不是不存在了？这是一个非常好的问题，答案是：对于有限体系来说情况确实会如此（除非有什么原因——比如对称性——禁止简并基态间的相互耦合）。但在量子场论中通常假定体系的空间体积趋于无穷，这时不同真空态之间的相互耦合趋于零，严格的对称性自发破缺只发生在这种情形下。

[16]戈德斯通定理也可以从几何上来理解。 $V(\varphi_a)$ ($a=1, \dots, N$) 可以看成是一个 N 维曲面，真空态对应于该曲面的一个极小值点，而该点处每一个独立的平坦方向（即二阶导数为零的方向）对应于一个无质量标量粒子。另一方面，每一个这种独立的平坦方向对应于一个可以使真

空态移到邻近点的连续对称变换。这种连续对称变换所表示的正是被真空态所破缺的对称性。这就表明无质量标量粒子与这种自发破缺的对称性一一对应。另外再补充一点：南部阳一郎曾在1960年提出过类似于戈德斯通定理的想法，但未引起足够重视。

[17]这里有一个很有意思的问题，那就是既然真正的对称性自发破缺是由量子有效势 V_{eff} 而非经典势函数 V 所决定的，那么在经典势函数 V 不具有简并真空态（从而不会产生对称性自发破缺）的情况下，是否有可能通过体现在有效势 V_{eff} 中的纯量子效应产生对称性自发破缺呢？答案是肯定的。如果哪位读者独立地想到了这个问题，那么祝贺你了，这说明你有非常敏锐的物理思维能力。如果你同时还具有第一流的理论基础，并且早生几十年的话，就有可能作出一个非常重大的理论发现，那便是1973年由美国物理学家科尔曼（Sidney Coleman, 1937—2007年）与温伯格（Erick Weinberg, 1947—）所发现的如今被称为科尔曼-温伯格机制（Coleman-Weinberg mechanism）的对称性破缺机制。

[18]一般来说，粒子物理学中的规范对称性指的就是“定域”规范对称性。不过在本节中，为突出“定域”所起的作用，我们有时会特意注明。

[19]用技术性的语言来说，在希格斯机制中对应于戈德斯通粒子的那些自由度可以被定域规范变换所消去（必须注意的是：“定域”二字在这里至关重要，整体的连续变换是不具有这种能力的）。从规范理论的角度讲，这相当于选取了一种被称为么正规规范（unitary gauge）的特殊规范。这种特殊规范的选取造成定域规范对称性的破缺，从而使原本受定域规范对称性所限必须无质量的规范粒子可以获得质量。人们有时把这种机制形象地描述为：规范粒子通过“吃掉”戈德斯通粒子而获得质量。另外要说明的是，这里所介绍的由希格斯等人提出的，被粒子物理标准模型所吸收的其实只是希格斯机制的一种最简单的实现形式——但似乎恰好就是自然界所采用的形式。

[20]电弱统一理论中的规范对称性破缺方式是 $SU(2) \times U(1)$ 破缺为 $U(1)$ ，由此产生的三个戈德斯通粒子通过希格斯机制使四个规范粒子中的三个（即 W^\pm 和 Z ）获得质量，剩下的一个（即光子）则维持了无质量。

[21]更确切地讲，标准模型中的汤川耦合是形如 $-\lambda \bar{\psi}_L \psi_R \phi + \text{h. c.}$ 的项，其中 ψ 为质量本征态（不同于弱本征态），L与R分别代表左右手征部分，h.c.代表厄密共轭。汤川耦合是费米子场与标量场之间唯一的可重整耦合。

[22]Gell-Mann将这一模型称为八正道（eightfold way），这一名称取意于佛教术语，所代表的是SU（3）分类模型中的八维表象。

[23] Ω^- 粒子于1964年被发现，它不仅量子数与理论预言完全一致，质量也非常接近理论的预期。

[24]当时盖尔曼是加州理工大学（California Institute of Technology）的教授，茨威格则是该校的研究生，他们虽在同一学校，但提出夸克模型是彼此独立的。夸克这一名称是盖尔曼所取，来自于爱尔兰作家乔伊斯（James Joyce, 1882—1941年）的小说《芬尼根的守灵夜》（*Finnegans Wake*）；茨威格提议的名字也很幽默，是“Aces”——即扑克牌中的“爱斯”。对茨威格来说，十分苦涩的经历是：同样标新立异的理论，Gell-Mann的文章应杂志编辑的亲自邀请发表在了欧洲核子中心（CERN）的新杂志《物理快报》（*Physics Letters*）上，而人微言轻的茨威格的文章却遭到拒稿而未能及时发表。茨威格后来转行离开了物理。

[25]这一点也适用于胶子或任何不处于色单态的粒子组合。不过要注意的是，它的严格数学证明是极其困难的。事实上，它是美国克莱数学研究所（Clay Mathematics Institute）悬赏百万美元征解的七大数学难题之一的“杨-米尔斯与质量隙”（Yang-Mills and Mass Gap）问题的一部分。不过许多物理学家对从数学上严格证明这一点并无太大兴趣，温伯格就曾经表示：“这一点肯定是正确的，因此我和其他一些人一样很乐意把证明留给数学家去做。”

[26]波利策等人因此而获得了2004年的诺贝尔物理学奖。比他们稍早，荷兰物理学家特·胡夫特也有过同样的发现，可惜没有发表。

[27]补充说明两点：（1）定义夸克质量所用的重整化方案（renormalization scheme）是 $\overline{\text{MS}}$ 。（2）夸克的“轻”和“重”是相对于量子色动力学中的特征能标 Λ_{QCD} （约为200~300MeV）来区

分的。

[28]有读者可能会问：既然有 $U(V)_V$ ，是不是也有 $U(1)_A$ ？在经典层次上答案是肯定的，但是在量子世界里， $U(1)_A$ 会被反常（anomaly）所破坏。

[29]感兴趣的读者请利用场量的宇称变换性质 $\phi(t, x) \rightarrow \gamma^0 \phi(t, -x)$ 自行证明 $V^{\mu a}$ 与 $A^{\mu a}$ 的变换性质 $V^{\mu a}(t, x) \rightarrow V^a_\mu(t, -x)$ 与 $A^{\mu a}(t, x) \rightarrow -A^a_\mu(t, -x)$ 。另外要注意的是，这里所说的矢量、轴矢量、标量、赝标量都是依据时空变换性质区分的，与那些量在 $SU(2)$ 内禀空间内的变换性质无关。

[30]由于s夸克也是轻夸克，因此我们的讨论可以扩展至包括s夸克，这是强子分类中存在 $SU(3)$ 近似对称性的原因——请注意这个 $SU(3)$ 是“味”对称性而不是“色”对称性。不过由于s夸克的质量较大， $SU(3)$ 对称性的近似程度远不如 $SU(2)$ 对称性来得高。

[31]在强子的命名中，有些带有质量参数， $\rho(770)$ 与 $a_1(1260)$ 就是两个例子。细心的读者可能要问：既然如此，这两个介子的质量怎么会是775MeV和1230MeV，而非770MeV和1260MeV呢？我把这个问题留给读者自己去思考。

[32]虽然从实验上观测到的强子谱来看，量子色动力学中的 $SU(2)_V \times SU(2)_A \times U(1)_V$ 对称性几乎肯定是破缺成了 $SU(2)_V \times U(1)_V$ （即手征对称性被破缺了），但这并不意味着量子色动力学的真空一定能够实现这一破缺方式。相反，能否实现这一破缺方式在很大程度上可以视为是对量子色动力学的检验。

[33] π 介子的质量远小于其他强子的质量，这一点很早就引起了人们的注意。为了解释这一现象，早在量子色动力学出现之前的1960年，南部阳一郎就提出可能存在一种极限情形（相当于后来的手征极限），在其中 π 介子是对称性自发破缺所产生的无质量粒子。中国物理学家周光召（1929—）也于1961年提出过类似的想法。

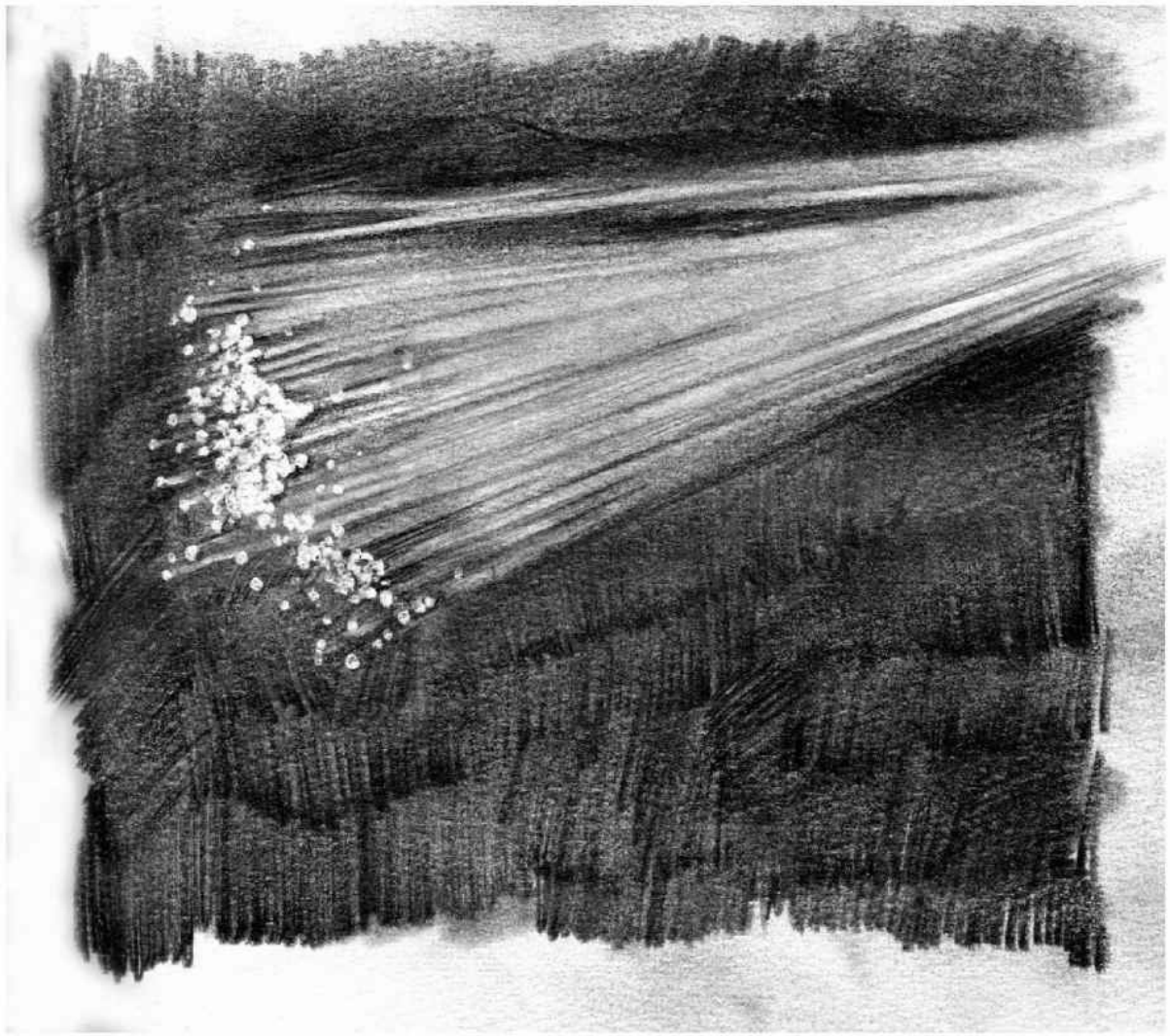
[34]不同的文献对 F_π 有不同的定义，彼此相差一个常数因子2或 $\sqrt{2}$

[35]这一结果在定性上是可以预期的，因为它大致等于量子色动力学中除夸克质量外的唯一能标 Λ_{QCD} 的三次方。感兴趣的读者可以（定性地）思考这样一个问题：在不考虑夸克质量的情况下，量子色动力学拉氏量中唯一的参数是无量纲的耦合常数，那么像 Λ_{QCD} 这样的能标是从何而来的？

[36]需要指出的是，对夸克质量的估计本身就在一定程度上运用了 π 介子（及其他几种介子）的质量。因此孤立地看，这里所谓的“吻合”带有循环论证的意味。但是人们对强子质量的计算是大量而系统的，涉及的粒子种类远远多于轻夸克的数目，当我们把所有这些计算综合起来看，这种“吻合”就不再是循环论证，而成为了很强的自洽性检验（consistency check）。这一点也适用于后文所述的对重子质量的计算。

[37]这些数值对比来自本文写作之初所参阅的文献，是大约十年前的研究结果。感兴趣的读者可以查阅一下新近文献，看是否有更好的结果。

[38]这个质量对应于一个由无质量的夸克和胶子组成的束缚态的质量。撇开计算上的复杂性不论，定性地讲，量子色动力学对这一质量的确定其实并不玄妙，它与量子力学对氢原子结合能的确定相类似——当然，氢原子在零质量极限下是不存在的。量子色动力学所具有的这种“质量隙”（mass gap）现象是高度非平凡的。另外，这个质量完全由相互作用所决定，在这一点上它有点类似于马赫早年的想法。只不过马赫设想的相互作用来自遥远的星体，而量子色动力学计算涉及的是微观世界的相互作用。感兴趣的读者可以思考一下：无质量的粒子为什么可以组成有质量的束缚态？



绘画：张京

纤维里的光和电路中的影^[1]

在一个周末的清晨，你上网查询了本市的景点信息，然后决定与家人一起参观新落成的科学博物馆；在博物馆里，你一边参观，一边用数码相机拍着像片；回家后，你用电子邮件将几张精选像片传给朋友，让他们分享你的快乐；晚上，你和家人围坐在一起，欣赏清晰的数字电视.....

你也许没有意识到，在这普通的一天里，你已反复成为了2009年诺贝尔物理学奖获奖成果的受益者。

2009年10月6日，拥有英国和美国双重国籍的华裔科学家高锟（Charles K. Kao），拥有加拿大和美国双重国籍的科学家博伊尔（Willard S. Boyle），以及美国科学家史密斯（George E. Smith）共同荣获了2009年的诺贝尔物理学奖^[2]。在这三人中，高锟“因光学通信中有关光在纤维中传输的突破性贡献”（for groundbreaking achievements concerning the transmission of light in fibers for optical communication）获得全部奖金（约140万美元）的一半，博伊尔和史密斯则“因发明一种成像半导体电路——CCD传感器”（for the invention of an imaging semiconductor circuit—the CCD sensor）而分享了另一半。

在本文中，我们将对这三位科学家的工作及其意义作一个简单介绍。

一、光纤，信息时代的大动脉

我们先来谈谈光纤。

简单地讲，光纤是一种能引导光在其中传输的纤维。初看起来，这并不是什么深奥莫测的东西，因为光——如我们早已知道——可在一切透明介质中传输，而光纤不过是制成纤维状的透明介质。这种用介质引导光的想法早在19世纪40年代初就已出现并付诸实验（所用介质是水和玻璃），它的一种早期应用是灯光喷泉（直到今天仍在用）。由于受光纤引导的光可以随光纤而弯曲，自20世纪20年代末起，人们开始设想用光纤来制作诸如胃窥镜之类的医学仪器，那些仪器可以深入患者体内，用光纤引导的光将患处的图像传输出来。

从物理上讲，光纤利用的是一种有趣的光学现象，那就是当光从折射率较高的介质（比如玻璃）射向折射率较低的介质（比如空气）时，在特定的角度范围内，入射光会在两种介质的交界面上被完全反射，而无法进入折射率较低的介质。这种现象被称为光的全内反射（total internal reflection），如图8所示。正是它保证了光纤内的光能够被光纤所引导，而无法轻易逃逸。

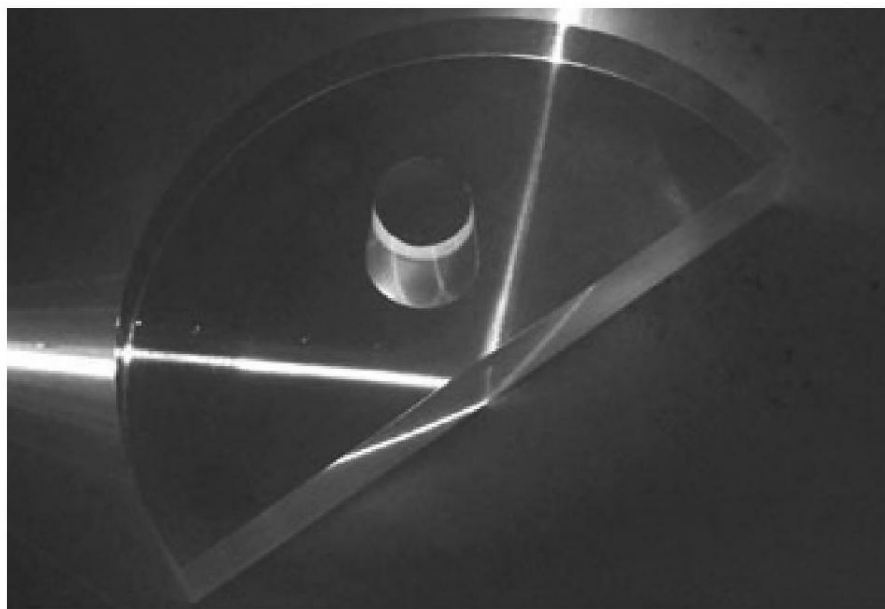


图8 光的全内反射

事情如果仅仅是这样，就没诺贝尔奖什么事了。人们在实际制作光纤时很快就发现，虽有全内反射在光纤的边界上把关，光纤中的光仍会迅速损耗。在20世纪60年代初，光在最好的光纤中，也只能传播区区20米就只剩下了1%左右。这使得光纤的应用只能局限于像医学仪器那样的短距离之内。

那么，光纤中光的快速损耗究竟是什么造成的呢？人们提出了一些可能的原因，比如光纤的弯曲，或光纤材料（比如二氧化硅）的晶体结构缺陷等。但是，任何实际应用中的光纤都不可能不弯曲，任何常温下的晶体结构也都不可能无缺陷。因此，若原因果真在这些方面，那光的快速损耗基本上就是“绝症”了。幸运的是，就在这光纤应用的整体前景面临极大挑战的时候，英国标准电信实验室（Standard Telecommunications Laboratories）的高锟与霍克汉姆（George Hockham）经研究发现^[3]，光的快速损耗并非上述原因所致，而主要是由于光纤中杂质——尤其是铁离子——对光的吸收与散射。他们这项研

究为光纤时代的降临开启了大门^[4]，因为既然罪魁祸首是杂质，我们要做的就只是对光纤材料进行提纯，而这是没有任何原则性困难的。

高锟等人的工作发表于1966年。4年之后，即1970年，美国玻璃制造商康宁公司就通过材料提纯，将原先20米的传输距离提升到了1000米^[5]。此后，就像所有技术领域的发展一样，这一纪录被一再刷新。自1975年起，英、美、日等国先后迈出了实用光纤通信的步伐。1988年，第一条跨大西洋的光纤电缆安装成功。现代的互联网、有线电视、电话通信等更是处处离不开光纤（图9）。可以毫不夸张地说，光纤已成为信息时代的大动脉。与传统的无线电通信相比，光纤所能传输的信息量要大得多，而且光纤所用之材料不仅蕴藏丰富，而且强度很高，具有得天独厚的应用优势。据估计，人们迄今铺设的光纤网络已达10亿千米左右，足可在地球与月亮之间绕一千多个来回。

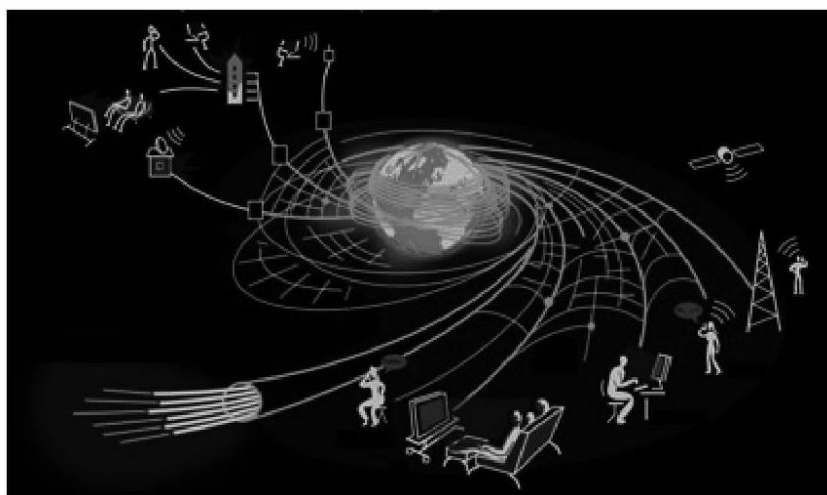


图9 光纤网络示意图

在光纤所传输的信息里，有很大一部分是数码影像，这些影像的由来将我们引向了今年诺贝尔物理学奖的第二项获奖工作：CCD。

二、CCD，数码摄影的电子眼

CCD是电荷耦合器件（charge-coupled device）的英文缩写。这种器件原本是作为一种电子内存而研发的。1969年秋天，美国贝尔实验室的博伊尔（Willard S. Boyle）和史密斯（George E. Smith）从事的就是这种研发工作。但CCD的真实用途几乎立刻就转变为了感光器件。

CCD的感光原理是建立在一种被称为光电效应（photoelectric effect）的现象之上的。这种现象曾被电磁波的发现者，德国物理学家赫兹（Heinrich Hertz）观察到——因此有时也被称为赫兹效应（Hertz effect），后来又经过了实验物理学家勒纳（Philipp Lenard）的研究，并由爱因斯坦利用当时还很新颖的光量子理论作出了理论解释（勒纳德与爱因斯坦因此分别获得了1905和1921年的诺贝尔物理学奖）。按照光电效应，适当频率的光照射到某些物质上时，会从物质中打出电子，其数目与光强成正比。

利用这一效应，博伊尔和史密斯将感光材料制成了一个由很多小单元组成的阵列^[6]，当光照射到阵列上时，会在每个小单元上打出一些电子。这些电子的数目分布很好地记录了入射光的强度分布。为了保存这些电子，博伊尔和史密斯让每个感光小单元都配有一个微小的电容。在感光过程结束后，这些小电容里的电子通过巧妙设计的电路逐排传递出去，并转变成为数字信号。这就是CCD的工作原理，而由那些数字信号组成的就是所谓的数码影像。由于CCD所用的将电子逐排传递出去的方式很像早年消防队员人工传递水桶的情形，因此这种器件也被称为“组桶式”器件（bucket brigade device），如图10所示。

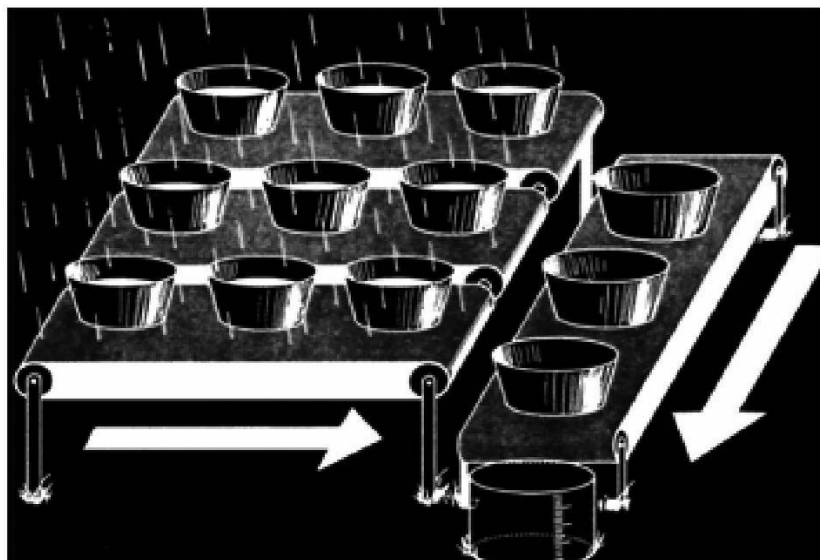


图10 CCD“组桶式”传输电子的比喻图

萌生CCD设想后的第二年，博伊尔和史密斯就将它用到了摄像机上；1972年，一家美国公司率先制造出了具有10 000（ 100×100 ）个感光单元的CCD传感器；1974年，第一张CCD天文相片问世；1975年，CCD摄像机达到了可用于电视转播的水准；1979年，CCD被首次安装到了天文望远镜上……CCD的发展走上了快车道。近年来，在CCD的冲击及其他因素的影响下，世界最大的胶卷生产商柯达公司（Eastman Kodak Company）陆续停止了普通胶片及胶片相机的生产。从某种意义上讲，这意味着一个时代——光学摄影时代——的终结。当然，它同时也是一个新时代——数码影像时代——日益成熟的标志。

那么，年轻的CCD与历史悠久的普通胶片相比究竟有什么优点呢？主要的优点有两个：一个是敏感度高，CCD能对90%左右的入射光子产生反应，也就是说，100个入射光子约有90个能在CCD的感光材料上产生电子，从而得到记录。而普通胶片及肉眼只能记录其中1~2个（高质量的胶片也只能记录10个左右）。另一个是适用范围广，CCD可用于从红外到X射线的各种波段。而普通胶片的适用范围却很狭窄，早期的普

通胶片甚至无法有效地涵盖可见光区内的红光，从而使得像褐矮星、红移值较高的类星体之类偏于长波的天体的发现大大延后。此外，普通胶片需要冲印，这对日常使用来说虽只是小麻烦，但对行星探测器来说可就要了命了，因为行星探测器大都是一去不复返的，不可能将胶片带回地球冲印。而CCD的数码信息却可以通过电波传回地球。我们今天看到的那些美轮美奂的行星图片，或哈勃太空望远镜（Hubble space telescope）拍摄的遥远星云都是因为有了CCD这只电子眼才成为了可能。对于观测天文学来说，CCD是一项能媲美望远镜与光谱仪的伟大发明。

光纤通信与CCD都是技术成就，但它们对于科学研究同样是必不可少的。今天的科学家们每天都在通过光纤大动脉交流着研究信息；翱翔在外层空间的太空望远镜每天都在用CCD电子眼窥视着这个让人着迷的宇宙。从这个意义上讲，获得今年诺贝尔物理学奖的虽是技术领域的工作，却对科学的发展有着意义深远的促进。

附录：获奖者小档案



高锟



博伊尔



史密斯

- 高锟（Charles K. Kao）：拥有英国和美国双重国籍的华裔科学家，1933年11月4日出生于中国上海，1965年获伦敦帝国大学（Imperial College London）电子工程学博士学位。高锟曾在英国标准电信实验室（Standard Telecommunications Laboratories）、香港中文大学等处任职，1996年退休，目前居住在美国。
- 博伊尔（Willard S. Boyle）：拥有加拿大和美国双重国籍的科学家，1924年8月19日出生于加拿大艾姆赫斯特（Amherst），1950年获加拿大麦吉尔大学（McGill University）物理学博士学位。博伊尔自1953年起在美国贝尔实验室（Bell Labs）任职，期间曾于20世纪60年代参与阿波罗登月计划，1979年退休，目前居住在加拿大。
- 史密斯（George E. Smith）：美国科学家，1930年5月10日出生于美国怀特普莱恩斯（White Plains），1959年获美国芝加哥大学物理学博士学位。史密斯自1959年起在美国贝尔实验室（Bell Labs）任

职，期间获得过几十项技术专利，1986年退休，目前居住在美国。

2009年10月11日写于纽约

[1]本文曾发表于《科学画报》2009年第11期（上海科学技术出版社出版）。

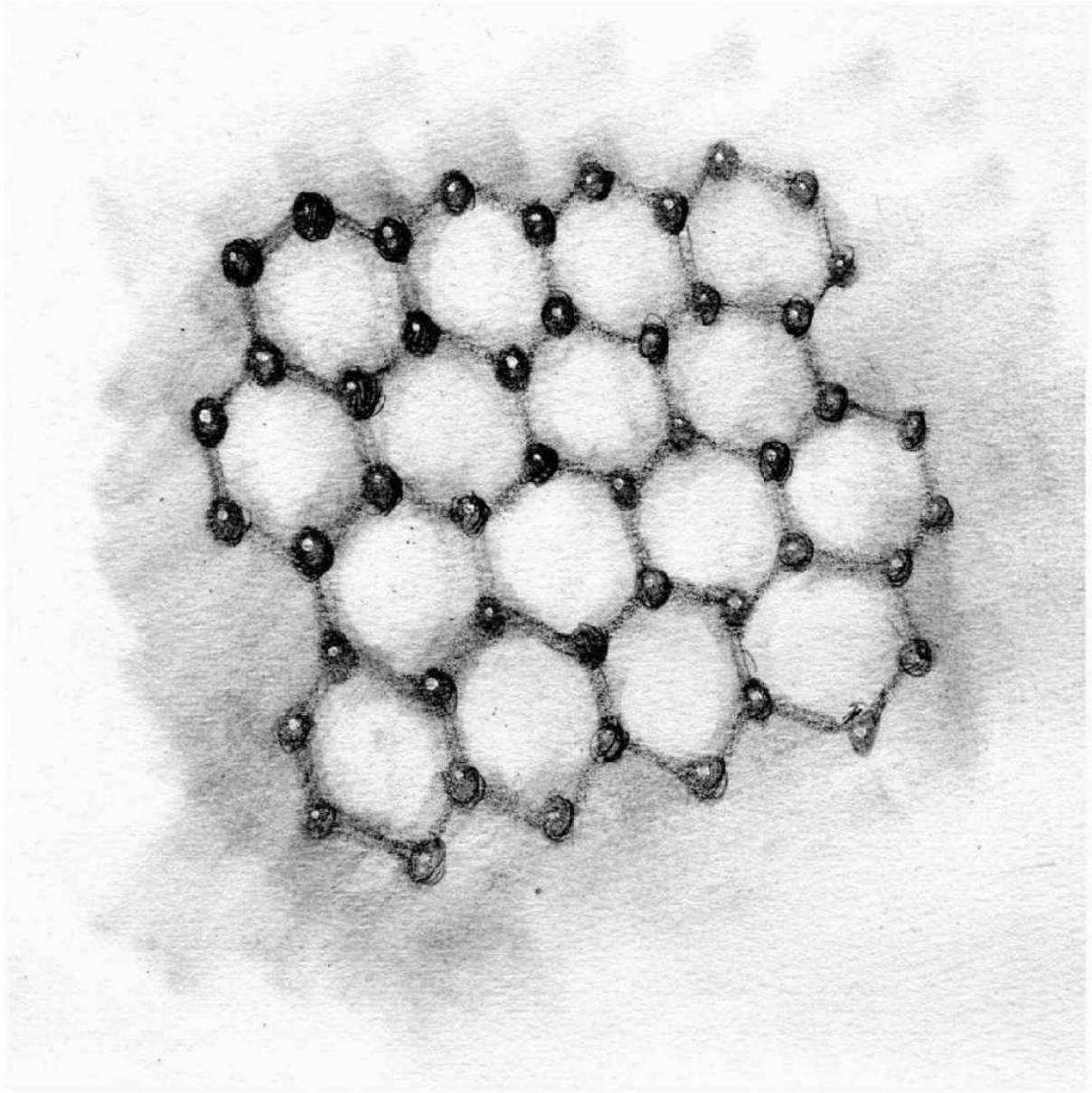
[2]由于这三位科学家的出生地及国籍丰富多彩，媒体在报道他们的获奖消息时充分发挥了灵活性。这三人在美国媒体上是三位美国科学家；在英国媒体上是一位英国科学家与两位美国科学家；在加拿大媒体上则是一位加拿大科学家与两位美国科学家。中国媒体自然也不落后，大陆媒体突出高锟的华人血统，香港媒体突出其任职香港中文大学的经历，台湾媒体则突出其“中央研究院”院士的身份。

[3]霍克汉姆于1969年获得电子工程学博士学位，一生获得过16项专利。高锟曾在2004年的一次访谈中提到，霍克汉姆从事的是理论研究。高锟成为当年那项研究的唯一获奖者，有可能是因为霍克汉姆当时还只是一位研究生。诺贝尔奖有过忽略研究生的先例，比如英国天文学家休伊什（Anthony Hewish）因脉冲星的发现而获得了1974年的诺贝尔物理学奖，他的学生贝尔（Jocelyn Bell Burnell）虽然是实际上的发现者，却没有获奖。当然，高锟在那篇论文发表之后又与其他人合作，对其他材料、其他波长的光纤应用进行了研究，为工业界指出了更具体的努力方向，这也很可能是他成为那项研究的唯一获奖者的原因。

[4]高锟被一些媒体称为“光纤之父”，不过“光纤之父”之名在此次诺贝尔物理学奖公布之前，通常是指美籍印度裔科学家卡潘尼（Narinder Kapany），他在20世纪50年代做过很多光纤方面的工作。另外要提到的是，与高锟的研究同年，德国科学家伯尔纳（M. Boerner）也提出了类似的观点，并在德、英、美等国获得了专利，不过此人不久后就去世了。

[5]用技术性的术语来说，康宁公司将光纤的损耗系数由每千米1000分贝减少为了17分贝。

[6]感光材料的选取标准是在所需的频率范围——比如可见光区——内具有显著的光电效应。



绘画：张京

石墨烯——从象牙塔到未来世界^[1]

2010年10月5日，瑞典皇家科学院（The Royal Swedish Academy of Sciences）宣布了2010年诺贝尔物理学奖的得主。荷兰籍俄裔物理学家盖姆（Andre Geim）和拥有俄罗斯及英国双重国籍的物理学家诺沃肖洛夫（Konstantin Novoselov）由于“对二维材料石墨烯的突破性实验”（for groundbreaking experiments regarding the two-dimensional material graphene）而共同荣获了这一奖项。

在本文中，我们将对这两位物理学家的获奖成果及其意义作一个简单介绍。

一、来自象牙塔的新材料

我们先来说明一下什么是石墨烯。这个名称中的“石墨”（graphite）二字我们大都不陌生，因为铅笔的笔芯就是由它和黏土混合而成的。从元素的角度讲，石墨是由碳元素组成的。在电子显微镜下，我们可以发现石墨的结构是层状的，每一层的碳原子都排列成紧密的蜂窝状六边形网格，层与层之间的距离则比较大，形成松散的堆砌^[2]（图11）。铅笔之所以在纸上轻轻一划就会留下痕迹，正是这种松散堆砌的结果。那么石墨烯（graphene）又是什么呢？它就是单层的石墨。

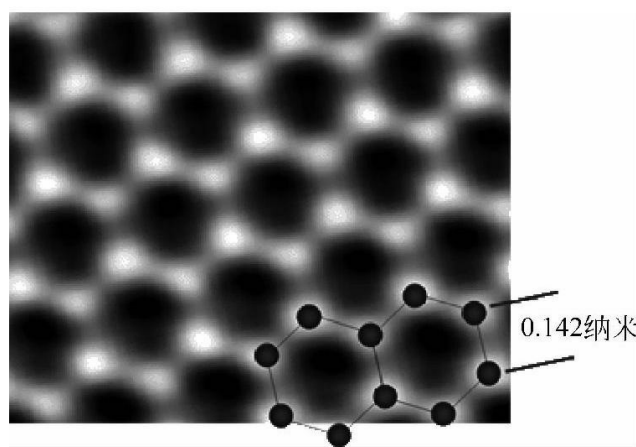


图11 电子显微镜下的石墨烯结构

石墨烯这个名称是从1987年开始使用的，但在那之前，就已经有人对这种单层原子组成的二维结构产生了兴趣，因为这种结构比现实世界里的三维结构来得简单，很适合当作例题收录在教科书里^[3]。通过这种象牙塔式的兴趣，人们开始对石墨烯的性质有了一些理论上的了解。这种了解，加上技术领域对新材料的需求日益旺盛，使人们对石墨烯产生了更现实的兴趣，试图将它由单纯的象牙塔物质“提拔”为真实材料。

初看起来，这种“提拔”似乎不会太困难。事实上，当我们用铅笔在纸上轻轻划过时，划痕中就可能会出现单层的石墨——即石墨烯。但问题是，铅笔的划痕从微观角度讲实在是太大了，在那里搜寻石墨烯简直就像是在整个喜马拉雅山脉中搜寻一片薄冰，即便找到也只能算是瞎猫碰上了死耗子。而科学家们需要的是系统的方法，是可以复制的成功，这却是相当困难的。直到21世纪初，人们所达到的最好业绩——即最薄的石墨片——也只能薄到几十层原子的水平。

更糟糕的是，有迹象表明，像石墨烯那样的二维材料有可能是注定只能存在于象牙塔里的。因为早在20世纪30年代，著名俄国物理学家朗道（Lev Landau）等人就已证明，二维材料的热运动涨落会破坏自身的结构。实验上制备石墨烯的种种失败尝试似乎也在佐证着这一结论，比如石墨层越薄，就越容易卷曲成球状或柱状，而无法维持平面结构^[4]。因此，制备石墨烯曾被很多人认为是注定无法成功的。

但以盖姆为核心的实验组却不信这个邪，决意尝试这一看似不可能的任务。这种尝试对他们来说，乃是一贯作风的延续。因为在盖姆的实验组里，对各种有趣、甚至有趣得近乎荒谬的事情的尝试已经达到了制度化的程度，他们每星期都几乎固定地拿出十分之一的时间来做一种所谓的“星期五之夜实验”（Friday evening experiment），专门尝试各种稀奇古怪的事情^[5]。制备石墨烯的工作也是从一个“星期五之夜实验”开始的。经过一些失败的尝试后，他们采用了所谓的“透明胶大法”（Scotch tape technique），即用透明胶粘住石墨层的两个面，然后撕开，使之分为两片。通过不断重复这一“大法”，并辅以其他手段，他们最终制备出了石墨烯^[6]。

盖姆和诺沃肖洛夫获奖后，许多媒体推出了渲染性的标题，比如《物理学家用透明胶和铅笔赢得诺贝尔奖》。这种标题容易给人一个错

觉，以为那是一项轻而易举的工作。事实上，盖姆实验组制备石墨烯的过程并不轻松，前后持续了一年多的时间，制备出的石墨烯则只有几平方微米，要用高倍显微镜才能观测。而且由于石墨烯是高度透明的，在观测及制备过程中还有一个如何分辨的问题。盖姆实验组解决这一问题的方法，是巧妙地利用了石墨烯在厚度300纳米的二氧化硅晶片衬底上产生的光线干涉效应。这一点是他们胜过其他研究组的关键所在。但即便如此，他们当时选用的衬底如果不是二氧化硅而是其他晶片，或者晶片的厚度不是300纳米，而是略大或略小，就都有可能无法分辨石墨烯。而他们当时之所以选用了恰到好处的衬底，据诺沃肖洛夫回忆乃是纯属偶然。因此，盖姆实验组的成功背后既有长时间的努力和巧妙的构思，也有运气的成分^[7]。当然，既然想到了正确的方法，发现合适的衬底应该是迟早的事情，从这点上讲，他们的成就并非偶然。

那么，这种辛辛苦苦制备出来的二维材料在我们这个三维世界里究竟有什么用处呢？在现实的用处出现之前，它在理论上的用处就已经吸引了科学家们的兴趣。物理学家们早在1956年就发现，托二维世界的福，石墨烯中的电子运动具有很奇特的性质，即电子的质量仿佛是不存在的^[8]。这种性质使石墨烯成为了一种罕见的可用于研究所谓相对论量子力学的凝聚态物质——因为无质量的粒子必须以光速运动，从而必须用相对论量子力学来描述。而更奇妙的是，那种相对论量子力学中的“光速”并不是真空中的光速，而只有后者的1/300。很多科学爱好者也许读过俄国物理学家伽莫夫（George Gamow）所写的科普作品《物理世界奇遇记》（*Mr. Tompkins in Paperback*），在那部作品中伽莫夫设想过一个光速很缓慢的世界。从某种意义上讲，石墨烯就是那样一个世界，它所具有的奇妙性质为理论物理学家们提供了一片研究相对论量子力学的新天地，使他们不仅可以把一些原先要用巨型加速器来研究的问题搬到自己的小型实验室里，而且还可以研究一些用巨型加速器都未曾

有机会透彻研究的东西，比如所谓的克莱因佯谬（Klein's paradox）或相对论量子力学特有的所谓“颤振”（zitterbewegung）效应，甚至还可以研究弯曲空间里的相对论量子力学——因为在石墨烯这个舞台上，弯曲空间不过就是弯曲的石墨烯而已。这些理论研究不仅题材新颖，而且还特别便于观测，因为石墨烯是二维的，所有现象都出现在表面上，不会像三维材料中的现象那样有可能跑到物质内部去。

除了成为研究相对论量子力学的新天地外，石墨烯还具有所谓的量子霍尔效应（quantum Hall effect），这种本身就是诺贝尔奖量级的重要效应以往是要在极低温下才能显现的，石墨烯却能将它带到室温下。诺沃肖洛夫在接受媒体采访时曾经表示，要让物理学家们改变自己的研究方向，必须用比他们所研究的有趣十倍的东西来引诱。石墨烯对很多理论物理学家来说看来就具有那样的魅力，因而吸引了众多的追随者。

二、通往未来世界的金桥

但石墨烯最吸引人的地方还在于它在现实世界里的可能应用。由于石墨烯的结构极为紧密和严整，哪怕在室温下都几乎没有任何缺陷，最大限度地发挥了众原子“集体的力量”，这使它不仅有比同等线度的钢铁还高两个数量级的强度，而且还有普通刚性材料难以企及的韧性，可以拉伸20%而不断裂。显示这种性质的流传最广的图片，是一幅猫躺在石墨烯制成的吊床上休息的想象图。这种由单层原子制成的吊床居然可以承受宏观物体的重量，无疑是令人惊叹的。那幅图片不够确切的地方，是没能显示出石墨烯的超薄特性。由于石墨烯的透光率高达97.7%^[9]，厚度却只有单层原子，因此如果真有那样的吊床，它不仅对于肉眼，甚至对于很多仪器都会是不可见的，我们看到的将是一只悬停在半空中的猫，就像《爱丽丝漫游奇境记》（*Alice's Adventure in Wonderland*）里那只柴郡猫（Cheshire cat）的笑容一样。

石墨烯如果只用来制作吊床，那显然是大材小用了。它更重要的可能应用是制成超薄、超轻、超强的材料，用于飞机、火箭、防弹衣等对材料性质要求极高的产品中。而它最能扣动人们想象之弦的可能应用，则是所谓的太空电梯。这种早在1895年就由火箭理论的先驱者、俄国科学家齐奥尔科夫斯基（Konstantin Tsiolkovsky）提出过的迷人设想，一直面临着一个致命问题，那就是找不到具有足够强度的材料来支撑线度达几万千米的巨型结构。石墨烯的出现使很多人重新燃起了希望。

除上述可能应用外，石墨烯的另一类可能应用则倚仗于它的电子运动性质。如我们在前面所述，石墨烯中的电子运动具有很奇特的性质，比如电子的质量仿佛是不存在的，而运动速度是所谓的“光速”。这些特

性，加上石墨烯结构在常温下的高度完美性，使得电子的传输及对外场的反应都超级迅速，几乎达到了人们梦寐以求的境界。体现到物理性质上，这使得石墨烯具有超常的导电性和导热性。这种性能既体现在纯净的石墨烯中，也可以部分地体现在含有石墨烯的复合材料中。而且更重要的是，石墨烯还可以用来制作晶体管，由于石墨烯结构的高度稳定性，这种晶体管在接近单个原子的线度上依然能稳定地工作。相比之下，目前勇挑大梁的以硅为材料的晶体管在10纳米（相当于几十层原子）左右的尺度上就会失去稳定性；而石墨烯中电子对外场的反应速度超快这一特点，又使得由它制成的晶体管可以达到极高的工作频率。事实上，IBM公司在2010年2月就已宣布将石墨烯晶体管的工作频率提高到了1000亿赫兹，超过了同等线度的硅晶体管^[10]。很多人相信，石墨烯将会成为硅的接班人，引领技术领域一个新的微缩时代的来临。

石墨烯的可能应用还有很多，比如它除了具有超高的强度和韧性外，还有不透水、不透气，以及抵御强酸、强碱的能力，这使它有可能成为制作保护膜的理想材料。而石墨烯既能导电又高度透明的特点，则使它有可能在制作液晶显示屏、触摸显示屏、太阳能电池板等领域大显身手。此外，用石墨烯制作的能快速充电的电池、容量超高的电容、能检测单个污染物分子的污染探测器、能用于量子计算机的特殊元件等，也都在构想或研制之中。

石墨烯从制备到获奖只用了短短六年的时间，与动辄要回溯几十年去“考古”的前几年的获奖成果相比，是非常快的。但在这六年里，由它开启的研究领域呈现了井喷的势头，几乎每个月都有新兴的研究方向被开辟出来。也许在不太遥远的将来，我们会开着由石墨烯电池驱动的车子去上班，在由石墨烯太阳能板提供能源的办公室里，用“内含石墨烯”（Graphene Inside——取代Intel Inside）的计算机从事工作。在假日

里——如果有闲钱的话——我们也许还可以乘坐用石墨烯材料建造的太空电梯去地球同步轨道欣赏地月同辉的奇景。这一切奇思妙想都得益于六年前的那项工作。在有关未来世界的构想中，很少有一种材料能像石墨烯那样大范围、跨领域地激发人们的想象力，并使人们因为看到实实在在的希望而有可能投入实实在在的努力。从这个意义上讲，它仿佛一座通往未来世界的金桥。

附录：获奖者小档案



盖姆



诺沃肖洛夫

- 盖姆（Andre Geim）：荷兰籍俄裔物理学家，1958年10月1日出生于俄国城市索契（Sochi），1987年获俄国科学院固体物理研究所博士学位。自1990年起，盖姆先后在英国诺丁汉大学（University of Nottingham）、丹麦哥本哈根大学（University of Copenhagen）、英国巴斯大学（University of Bath）、荷兰内梅亨大学（Radboud University Nijmegen）等地工作过。2001年，盖姆成为英国曼彻斯特大学（University of Manchester）物理学教授，并于2002年起担任曼彻斯特介观科学及纳米技术中心（Manchester Centre for Mesoscience and Nanotechnology）主任。
- 诺沃肖洛夫（Konstantin Novoselov）：拥有俄罗斯及英国双重国籍的物理学家，1974年8月23日出生于俄国城市尼茨塔吉尔（NizhnyTagil），2004年获荷兰内梅亨大学博士学位。诺沃肖洛夫是盖姆的学生及长期合作者，自2001年起，与盖姆一起在英国曼彻斯特大学工作。诺沃肖洛夫是自1973年以来最年轻的诺贝尔物理学

奖得主。

2010年10月11日写于纽约

[1]本文曾发表于《科学画报》2010年第11期（上海科学技术出版社出版）。

[2]石墨每一层上的碳原子间距约为0.142纳米，层与层的间距则为0.335纳米，后者是依靠微弱的范德瓦耳斯力（van der Waals force）结合起来的，因而是松散的堆砌。

[3]当然，这里所谓的“二维”不是几何上的二维，而仅仅是指垂直方向上的物理自由度可以忽略的情形。

[4]不过那种球状或柱状的结构对于石墨烯的制备来说虽是“麻烦制造者”，本身却都是绝顶的好东西：前者是所谓的富勒烯（fullerene），它的发现者获得了1996年的诺贝尔化学奖；后者则是大名鼎鼎的纳米管（nanotube），也是一种令人着迷的新材料。

[5]盖姆曾经因为在这种“星期五之夜实验”中进行过“磁悬浮青蛙”实验，而获得了2000年的搞笑诺贝尔物理学奖（Ig Nobel Prize in Physics）。他是迄今唯一一位同时获得过搞笑诺贝尔奖和诺贝尔奖的人。

[6]有读者可能会问：既然朗道曾经证明过二维材料的涨落会破坏物质结构，怎么还可能制备出石墨烯呢？答案是，朗道的证明是针对大面积（理论上是无穷大）的体系的，而人们最初制备的石墨烯只有几平方微米。另一方面，朗道的证明考虑的是严格的平面，而真实的石墨烯会在三维空间里波动，从而耗散掉一部分涨落能量。因此石墨烯的出现虽然出人意料，却不是不可理解的。

[7]制备石墨烯——尤其是大样品——的难度还可以从另一个角度来印证，那就是石墨烯的价格。直到2008年4月，石墨烯的价格依然高到令人瞠目的每平方厘米一亿美元，堪称史上最贵的材料。不过最近两年，人们制备石墨烯的能力已突飞猛进，最大样品的线度已超过70厘米，价格也已暴跌（因此千万不要囤积石墨烯，它很重要，但绝不可能使你发财）。

[8]确切地说，那并非电子，而是电子与石墨烯晶格相互作用所产生的准粒子（quasiparticle），是石墨烯的低能激发态。

[9]石墨烯的这个透光率（对应于吸收率2.3%）是一个漂亮的理论结果，精确公式为 $(1+\pi\alpha/2)^{-2}$ ，其中 α （ $\approx 1/137$ ）是所谓的精细结构常数。很多媒体引用的是这一公式的近似式： $1-\pi\alpha$ 。

[10]IBM所宣称的1000亿赫兹其实是“适度浮夸”的结果，实际试验中所达到的频率约为300亿赫兹。

囚禁的量子，开放的应用^[1]

2012年10月9日，一位68岁的法国老人与妻子在街头散步，当他们路过一条街边的长椅时，电话忽然响起，老人被告知获得了诺贝尔物理学奖。同样被“搅扰”的还有大西洋彼岸的一位也是68岁的美国老人，电话响起时他还在睡梦中，但无论什么梦也没有电话里的消息更美：他也获得了诺贝尔物理学奖。

这两位天各一方，但恰巧同岁的老人分别是法国物理学家阿罗什（Serge Haroche）和美国物理学家维因兰德（David Wineland），之所以获奖，是因为他们实现了“使得对单个量子体系的测量与操控成为可能的突破性实验方法”（for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems）。他们将共同分享崇高的荣誉，以及虽因金融危机而缩水，但数量依然可观的800万瑞典克朗（约合110万美元）的奖金^[2]。

在本文中，我们将对这两位物理学家的工作及其意义作一个简单介绍。

一、小有小的麻烦

美国物理学家费恩曼曾以一个有趣的问题作为《费恩曼物理学讲义》（*The Feynman Lectures on Physics*）的开篇，那就是：假如因为某种灾变，在所有科学知识中只有一句话能传之于后代，什么话能用最少的文字包含最多的信息？费恩曼认为，那应该是所谓的“原子假设”，即所有物质都是由原子组成的^[3]。不过，这句话包含的信息虽多，要想破译却并不容易。事实上，早在两千多年前的古希腊就有先贤猜测过物质是由原子组成的（“原子”一词的英文atom就来自希腊文 $\alpha\tau\omicron\mu\omicron\varsigma$ ，含义为“不可分割的”），但直到18世纪才开始有了现代意义下的原子理论，而原子的真正奥秘，则直到20世纪才开始揭晓。

为什么呢？因为原子实在太小了，既看不见，也摸不着。

如今我们知道，原子并非是“不可分割的”，它由更基本的粒子所组成，并且与那些粒子一样，遵守一种被称为量子力学（quantum mechanics）的奇妙规律。这种规律与我们习以为常的宏观世界的规律完全不同，在发现之初曾带给物理学家们极大的震动。直到很多年后，当那种规律逐渐褪去新鲜的外衣，甚至已变成物理系学生的常识时，想在最直接的意义上体验它们仍是极为困难的事情。

为什么呢？依然是因为原子实在太小了，既看不见，也摸不着。

由于这一原因，物理学家们对原子——或者更一般的，对量子体系——的很多观测都不是针对单个原子（或量子体系）的。比如他们观测的原子光谱乃是由很多原子共同发射的。而在有条件观测单个原子（或量子体系）的实验中，由于观测对象太小，往往观测一结束，观测对象

本身也就“人间蒸发”或“香消玉殒”了，比如用云室或气泡室（这两者的发明者分别获得了1927年和1960年的诺贝尔物理学奖）观测粒子，或用照相设备观测光子就都是如此。

那么，有没有什么办法，能够观测甚至操控单个量子体系，同时还让它继续存在（从而还可以继续观测或操控）呢？维因兰德和阿罗什——在他们各自同事的鼎力合作下——所解决的正是这个问题。他们凭借高超的实验技巧，将单个量子体系囚禁起来，然后用细微而巧妙的“探针”去观测甚至操控它，从而完成了近乎“不可能任务”（mission impossible）的壮举，为上述问题提供了肯定答案（图12）。

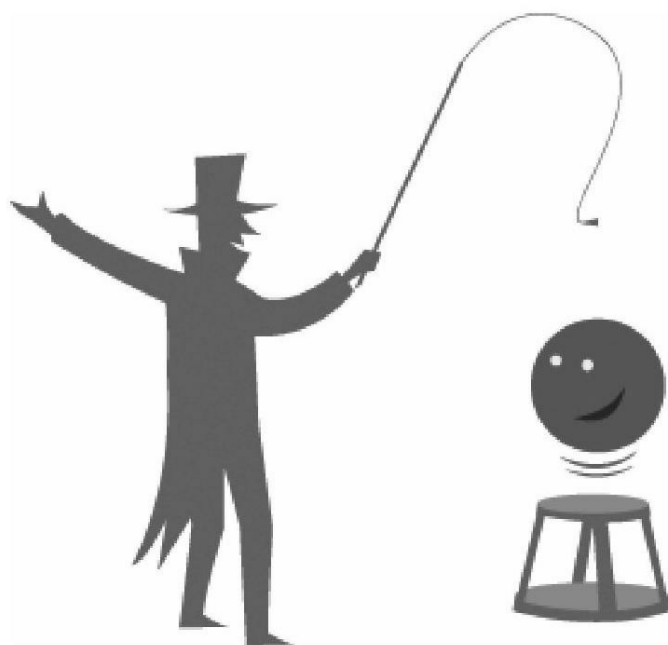


图12 维因兰德和阿罗什完成了近乎“不可能任务”的壮举

下面我们就对他们的方法做一个简单介绍。

二、囚禁的量子

维因兰德采用的方法是将单个的离子（离子是失去或得到若干电子——从而带电——的原子），比如铍离子 Be^+ （它是失去一个电子的铍原子），利用其带电的特征，囚禁在用电磁场组成的“牢笼”中，然后以光子作“探针”去探测和操控它。这话说起来简单，实现起来却极不容易，单是那“牢笼”——它的“学名”叫做离子阱（ion trap）——本身就已经是一个诺贝尔奖级别的成就（它的实现者获得了1989年的诺贝尔物理学奖）^[4]。为了确保被囚禁的是单个（或少数几个）离子，还需要辅以超高真空（以便排除其他粒子的干扰）和超低温（以便排除热运动的干扰）等技术。其中后者采用的乃是维因兰德与同事亲自参与研发的绝活：边带冷却技术（sideband cooling）^[5]。当这些极不简单的配置完成之后，维因兰德又通过激光脉冲（光子），将被囚禁离子的内部状态（即电子能态）叠加起来。这种状态叠加是量子力学有别于经典物理的奇妙特征，科普读物中常见的诸如“粒子既在这里，又在那里”，“猫既是死的，又是活的”，等等吸引眼球的表述都源自于此。但维因兰德能做到的还不止这些，通过对激光脉冲的巧妙选择，他还可以对状态叠加的方式进行操控，比如将离子内部状态的叠加转变为外部状态（即离子在“牢笼”内的振动状态）的叠加，甚至将一个离子的状态叠加转变为另一个离子的状态叠加。

与维因兰德的方法几乎恰好相反，阿罗什的囚禁物是被维因兰德当作“探针”的光子，而“探针”则类似于维因兰德的囚禁物，是一种被称为里德堡原子（Rydberg atom）的特殊原子，它的电子处于很高的能态上，从而使整个原子“发胖”到惊人的程度。比如阿罗什所用的铷（Rb）原子就“发胖”到了普通铷原子的500倍左右^[6]。在阿罗什的方法中，囚

禁光子所用的是以超导材料铌（Nb）制作的一对相距2.7厘米的球面镜，这对球面镜的工艺极为高超，构成了一个反射性质近乎完美的空腔（cavity）。光子在其中可以被反射十几亿次而不被吸收（在这过程中走过的总距离可以绕地球一圈）。在这些同样极不简单的配置完成之后，阿罗什又通过特殊空腔中的电磁波，使作为“探针”的里德堡原子处于两个电子能态的叠加之中，并使之以可控制的速度穿越囚禁了光子的空腔。在这里，阿罗什做了另一个巧妙安排，使被囚禁光子的能量与里德堡原子所能吸收的能量稍稍错开，从而保证光子不会被里德堡原子所吸收（别忘了，这一整套方法的使命之一就是保障量子体系继续存在）。而更巧妙的是，尽管光子不会被吸收，它与里德堡原子的相互作用仍能对后者产生影响，改变后者那两个叠加能态间的相位。这样，阿罗什就可以通过研究穿越后的里德堡原子那两个叠加能态间的相位，而获得有关被囚禁光子的某些信息（比如光子的数目）。

上述两种方法的实现无疑都需要极高超的技术。不过，此类“工艺性”的工作要想获得诺贝尔奖，通常还需满足一个额外条件，那就是具有应用价值。此次获奖的工作很好地符合了这一条件，因为其所实现的“使得对单个量子体系的测量与操控成为可能的突破性实验方法”在理论与实用上都有着重要应用。

三、开放的应用

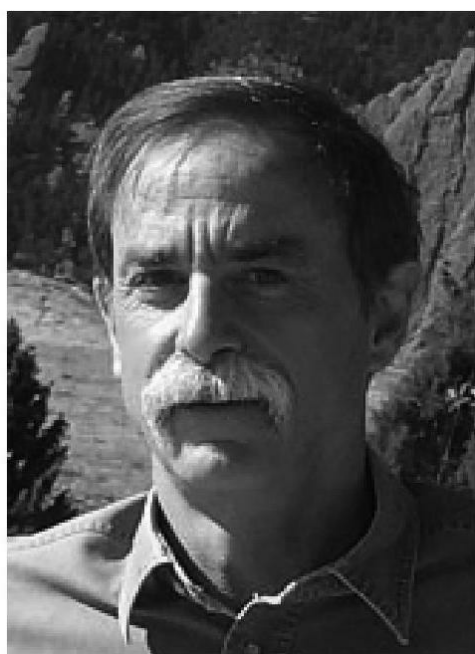
在理论上，对一个量子体系进行观测或操控，同时还让它继续存在，使得人们设计出了一些巧妙的实验，来观测量子体系状态演变的过程（以往的实验由于是“一锤子买卖”，对被观测体系具有“毁灭性”，从而无法做到这一点），甚至观测使一些物理学家深感困惑的量子体系的状态因为与外部环境的相互作用而往经典状态过渡的过程，其中包括对大名鼎鼎的“薛定谔的猫”（Schrödinger's cat）的生死过程的观测^[7]。那样的实验已经有人做了。比如阿罗什本人的研究组就于2008年做了那样的实验，甚至将观测到的量子状态往经典状态过渡的过程制成了“影片”。

在实用上，此次获奖工作最引人注目的应用是在量子计算机领域。这是近年来被讨论得很多的领域，在乐观者看来，量子计算机若成为现实，对社会的变革将不亚于如今的计算机在过去几十年所带来的变革。不过，量子计算机的理论虽然美丽，面临的技术困难却极为巨大，其中一个很大的困难就是作为核心元件的量子体系必须能单个地、不受破坏地被测量与操控，而且各个量子体系的状态还必须能相互传递（就像经典计算机必须能在各元件间传递信息一样）。这个困难在过去几乎是难以克服的，此次的获奖工作却为之带来了曙光，比如维因兰德所实现的对状态叠加的操控，以及状态叠加在不同离子间的相互转变，就正是克服上述困难所需要的技术。这一点维因兰德本人也看得很清楚——事实上，他的研究组早已展开了这方面的探索，甚至在一定程度上构造出了量子计算机的雏形，实现了最简单的逻辑运算。一些其他实验组也正在积极努力之中。当然，这一切距离真正有实用价值的量子计算机还相差很远。

此次获奖工作的另一项很有价值的应用是建造超高精度的新型时钟。这一应用虽不像量子计算机那样富有未来色彩，所取得的进展却要扎实得多。维因兰德所供职的美国国家标准技术研究所正是这方面的“领头羊”。在这一应用中，用维因兰德所实现的方法囚禁起来的工作频率（即作为计时基础的两个能级之间的量子跃迁的频率）在光学波段的离子取代了传统原子钟所采用的工作频率在微波波段的铯（Cs）原子。目前，这种新型时钟已经达到了比传统铯原子钟高两个数量级的精度。在那样的精度下，哪怕从宇宙大爆炸之初开始计时，迄今的累计误差也只有区区几秒。

这些或已成为现实，或仍处于开放的想象空间里的应用，使此次的获奖工作有可能对未来科学与技术的发展产生深远影响。

附录：获奖者小档案



维因兰德



阿罗什

- 维因兰德（David Wineland）：美国物理学家，1944年2月24日出生于美国威斯康星州的密尔沃基（Milwaukee），1970年获哈佛大学（Harvard University）物理学博士学位，目前在美国科罗拉多州的国家标准技术研究所（National Institute of Standards and Technology）任职。维因兰德的主要研究方向为量子光学（quantum optics）及其应用。
- 阿罗什（Serge Haroche）：法国物理学家，1944年9月11日出生于当时受法国控制的摩洛哥城市卡萨布兰卡（Casablanca），1971年获巴黎第六大学（Université Pierre et Marie Curie）的物理学博士学位，目前在法国巴黎（Paris）的法兰西公学院（Collège de France）任教。阿罗什的主要研究方向为量子光学及其应用。

2012年10月11日写于纽约

[1]本文曾发表于《科学画报》2012年第11期（上海科学技术出版社出版）。

[2]在过去若干年里，每个奖项的奖金为1000万瑞典克朗。

[3]这里我们稍稍偷了点懒，费恩曼想要传给后代的话还包括了原子处于永恒的运动之中，以及它们太过靠近时彼此排斥，稍稍远离时彼此吸引这几点。

[4]确切地说，最常用的离子阱有两种，一种叫做彭宁阱（Penning trap），另一种叫做保罗阱（Paul trap），他们的实现者分享了1989年的诺贝尔物理学奖。维因兰德所使用的是保罗阱。

[5]边带冷却技术简单地说，是用能量为 $\omega_i - \omega_v$ （其中 ω_i 为离子的内部能级差， ω_v 为离子在“牢笼”内的振动能级差）的光子，将处于振动能级 $n > 0$ 的离子激发到内部能级更高，但振动能级只有 $n - 1$ 的状态上（因为那样的光子只能将离子激发到那样的状态），然后让离子自行跃回原先的内部能级。由于离子在跃回过程中会优先维持振动能级不变，因此过程终了时离子的内部能级不变，振动能级却降为了 $n - 1$ 。重复这一过程（在必要时针对所需要的内部能级差调整光子能量），可以使振动能级最终降为基态 $n = 0$ ，从而达到冷却的目的。

[6]阿罗什所用的“发胖”后的铷原子的线度约为125纳米（nm），而普通铷原子的线度约为0.25纳米。

[7]当然，这是夸张的说法，事实上那猫被“掉包”成了一个量子体系，从而偏离了薛定谔拿猫“开涮”的本意——即通过引进作为宏观客体的猫，而彰显量子测量过程的佯谬性。不过包括诺贝尔委员会（Nobel Committee）提供的获奖作品介绍在内的大量资料和报道都已迫不及待地引入了“薛定谔的猫”一词。作为科普，我们姑且“从众”，但在这里略做说明，以图确切。

第三部分 星际旅行漫谈

因为星星在那里

Space, the final frontier!

StarTrek: The Next Generation

试图挑战自然的人常会被问到为什么要用自己的生命去冒险。我有一位酷爱登山的朋友，一同在哥伦比亚大学（Columbia University）念研究生期间的某个夏天，他登上了北美洲的最高峰——海拔6 194米的麦金利峰（Mount McKinley）。我在系里遇见了刚从雪域高原回来的他。锐利的紫外线灼黑了他的皮肤，使我几乎认不出来，但一种敬意在我心中油然而生。我没有问他为什么要去登山，我知道登山家有一句震撼人心的名言：因为山在那里（Because it's there）。

小时候喜欢看星星，常可以看上几个小时不知倦怠。我知道天空中几乎每一颗小小的星星都要比我们脚下这个看似巨大的蓝色星球大上数百万倍，“大”与“小”竟以如此瑰丽的方式相互嵌套，那是何等的深邃和奇异啊！

30年前的1972年，人类向外太阳系发射了名为“先驱者10号”（Pioneer 10）的行星探测器。一年后又发射了它的姊妹探测器“先驱者11号”（Pioneer 11）。它们已先后飞出了我们的太阳系（如果以冥王星轨道作为太阳系边界的话）。目前“先驱者10号”大约在距地球120亿

千米之外，正向着65光年外的金牛座（Taurus）的毕宿五（Aldebaran）星飞去，以目前的速度计算将在约200万年后抵达。“先驱者11号”则将在约400万年后掠过天鹰座（Aquila）的一颗恒星。

200万年对人类来说是一段过于漫长的时间：200万年前人类还过着茹毛饮血的穴居生活；200万年后当“先驱者10号”迎来自己孤独航程中第一缕耀眼的异星光芒时，人类也许早已在愚昧的战乱中成为了无言的化石。

登山家面对的是以人类微薄的体力去挑战大自然的伟岸，星际旅行家面对的则是以人类短暂的生命去跨越星际间几乎无限的距离。人类的平均寿命在过去几十年间虽然有所增长，但自然衰老依然是无可抗拒的规律。即使在基因图谱逐渐被揭开的今天，也没有迹象表明人类的寿命会在可预见的将来获得数量级上的延长。

从逻辑上讲，要让星际旅行家用短暂的生命去跨越近乎无限的时空，不外乎有两类方案：一类是从星际旅行家本身入手，设法在各种意义下延长其生命；另一类是从时空入手，设法利用或改变其结构，达到缩短空间距离或突破速度极限的目的。具体地讲，常见的设想有以下几种：

- 从星际旅行家本身入手的方案：
 - 用极低温“冷冻”的方法延长生命。
 - 用巨型空间站代替飞船，以群体繁衍的生命取代个体的生命。
 - 建造飞行速度接近光速的飞船，利用相对论的时间延缓效应达到延长生命的目的。
 - 将星际旅行家分解为基本粒子流或信息流以光速或接近光速的速度传播，并在目的地复现乘员。

- 从时空入手的方案：
 - 通过“虫洞”（wormhole）实现时空间的“捷径”（short-cut）旅行。
 - 通过“曲速引擎”（warp drive）实现“超光速”旅行。

“星际旅行漫谈”这个系列的文章将以目前所知的物理学规律为依据，来讨论其中的若干种方案，无论它们是出自科学家、工程师还是科幻小说家之手。

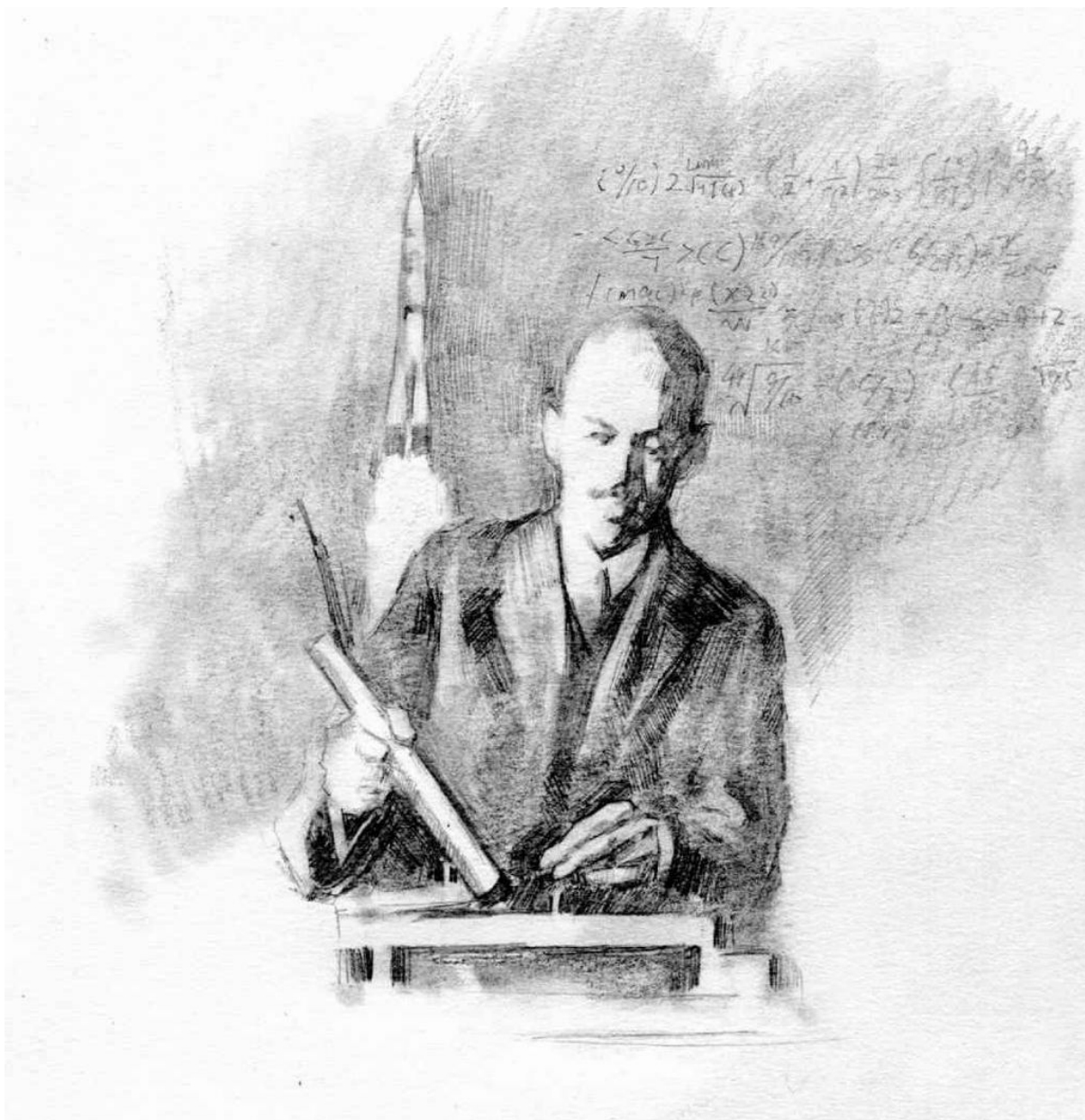
这些方案是人类探索璀璨星空的梦想的延续。

自远古以来这种梦想就以这样那样的方式存在着，历经无数的磨难和挫折，却从来不曾消失过。

因为人类的好奇心不可磨灭，因为星星在那里。

2002年7月24日写于纽约

纪念“先驱者10号”发射30周年



绘画：张京

火箭：宇航时代的开拓者

一、引言

这个星际旅行系列原本是为了讨论未来的星际旅行技术而写的。不过今天却要来讨论一种比较“土”的技术：火箭。之所以讨论火箭，主要的原因有两个：一个是因为我国的第一艘载人飞船“神舟五号”即将发射^[1]，在这个中国宇航员即将叩开星际旅行之门的时刻，我们这个系列不应缺席，也不应让火箭这位宇航时代劳苦功高的开拓者在这个系列中缺席。另一个是因为火箭虽然是一种不那么“未来”的技术，但在我和读者诸君能够看得到的未来，承载人类星际旅行之梦的技术很有可能仍然是火箭这匹识途的老马。

二、宇宙速度

火箭理论的先驱、俄国科学家齐奥尔科夫斯基（Konstantin Tsiolkovsky, 1857—1935）有一句名言：“地球是人类的摇篮。但人类不会永远躺在摇篮里，他们会不断探索新的天体和空间。人类首先将小心翼翼地穿过大气层，然后再去征服太阳周围的整个空间。”

星际旅行是一条漫长而坎坷的征途，人类迄今在这征途上所走过的部分几乎恰好就是“征服太阳周围的整个空间”，而这征途上的第一站也正是“穿过大气层”^[2]。

在人类发射的航天器中，数量最多的就是那些刚刚“穿过大气层”的航天器——人造地球卫星，迄今已发射了数以千计。其中第一颗是1957年10月4日从苏联的拜克努尔航天发射场（Baikonur Cosmodrome）发射升空的“卫星一号”（Sputnik 1）。

从运动学上讲，这些人造地球卫星的飞行轨迹与我们随手抛掷的一块石头的飞行轨迹是属于同一类型的。我们抛掷石头时，抛掷得越快，石头飞得就越远，石头飞行轨迹的弯曲程度也就越小。倘若石头抛掷得如此之快，以致于飞行轨迹的弯曲程度与地球表面的弯曲程度相同，石头就永远也不会落到地面了^[3]。这样的石头就变成了一颗环绕地球运转的小卫星，这一点早在牛顿（Isaac Newton, 1642—1727）的《自然哲学的数学原理》（*Mathematical Principles of Natural Philosophy*）中就有过精彩的图示（图13）。一般地讲，石头也好，卫星也罢，它们的飞行轨迹都是椭圆^[4]。对于石头来说，如果抛掷得不够快，那它很快就会落到地面，从而我们就只能看到椭圆轨道的一个极小的部分，那样的部分近似于一段抛物线（感兴趣的读者请自行证明这一点）。

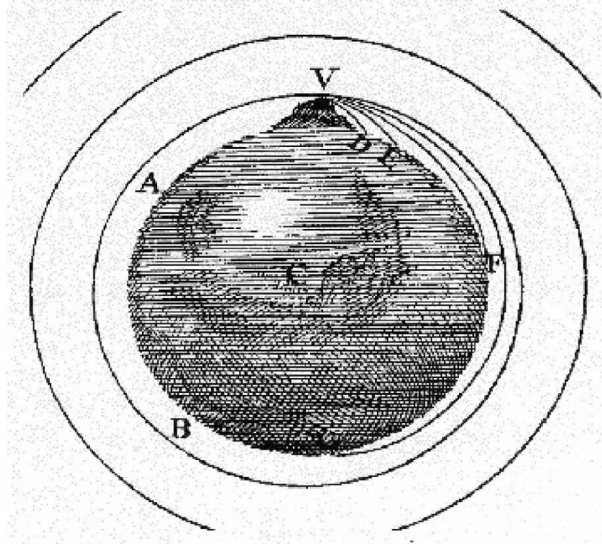


图13 牛顿《自然哲学的数学原理》的插图

那么，一块石头要抛掷得多快才能不落回地面呢？或者说一枚火箭要能达到什么样的速度才能发射人造地球卫星呢？这个问题的答案很简单——尤其是对于圆轨道的情形。在圆轨道情形下，假如轨道的半径为 r ，卫星的飞行速度为 v ^[5]，则维持卫星飞行所需的向心力为 $F=mv^2/r$ （ m 为卫星质量），这一向心力来源于地球对卫星的引力，其大小为 $F=GMm/r$ （ M 为地球质量）。由此可以得到 $v=(GM/r)^{1/2}$ 。假如卫星轨道很低（即轨道离地球表面很近），则 r 约等于地球半径 R ，由此可得 $v\approx 7.9$ 千米/秒。这个速度被称为“第一宇宙速度”（first cosmic velocity），它是人类迈向星空所要达到的最低速度。

不过，细心的读者可能会从上面的计算结果中提出一个问题，那就是 $v=(GM/r)^{1/2}$ 随着轨道半径的增加反而在减小，这说明轨道越高的卫星飞行速度越小。但是直觉上，把东西扔得越高难道不应该越困难吗？再说，倘若把卫星发射得越高所需的速度反而越小，那么 $v\approx 7.9$ 千米/秒这个“第一宇宙速度”岂不就不再是发射人造地球卫星所要达到的最低速度了？这些问题的出现，表明对于发射卫星来说，卫星的飞行速度并不

是所需考虑的唯一因素。那么，还有什么因素需要考虑呢？答案是很多，其中最重要的一个是引力势能。事实上描述发射卫星困难程度的更有价值的物理量不是卫星的飞行速度，而是发射所需的能量，也就是把卫星从地面上的静止状态送到轨道上的运动状态所需提供的能量。因此我们改从这个角度来分析。在地面上，卫星的动能为零^[6]，势能为 $-GMm/R$ ，总能量为 $-GMm/R$ ；在轨道上，卫星的动能为 $mv^2/2=GMm/2r$ （这里运用了前面得到的 $v=(GM/r)^{1/2}$ ），势能为 $-GMm/r$ ，总能量为 $-GMm/2r$ 。因此发射卫星所需的能量为 $GMm/R - GMm/2r$ 。这一能量相当于把卫星加速到 $v=[GM(2/R - 1/r)]^{1/2}$ 所需的能量。由于 $r>R$ ，这一速度显然大于 $v=(GM/r)^{1/2}\approx 7.9$ 千米/秒（而且也符合轨道越高发射所需能量越多这一“直觉”）。这表明“第一宇宙速度”的确是发射人造地球卫星所需的最低速度，只不过它表示的并不是卫星的飞行速度，而是火箭提供给卫星的能量所对应的等价速度。在发射卫星的全过程中，火箭本身的飞行速度完全可以在任何时刻都低于这一速度。

上面的分析是针对圆轨道的，那么椭圆轨道的情况如何呢？在椭圆轨道上，卫星的飞行速度不是恒定的，分析起来要困难一些，但结果却同样很简单，卫星在椭圆轨道上的总能量仍然为 $-GMm/2r$ ，只不过这里 r 表示所谓的“半长径”，即椭圆轨道长轴长度的一半。因此上面关于“第一宇宙速度”是发射人造地球卫星所需的最小（等价）速度的结论对于椭圆轨道也成立，是一个普遍的结论。

在人造地球卫星之后，下一步当然就是要把航天器发射到更远的地方——比方说月球上。为了实现这一步，火箭需要达到的速度又是多少呢？这个问题的答案也很简单，不过在回答之前先要对“更远的地方”做一个界定。所谓“更远的地方”，指的是离地心的距离远比地球半径（约

为 6.4×10^3 千米)大,但又远比地球与太阳之间的距离(约为 1.5×10^8 千米)小。之所以要有后面这一限制,是因为在讨论中我们要忽略太阳的引力场^[7]。由于航天器离地心的距离远比地球半径大,因此与发射前在地面上的引力势能相比,它在发射后的引力势能可以被忽略;另一方面,由于航天器不再做环绕地球的运动,其动能也就不再受到限制,最小可能的动能为零。(请读者想一想,这一动能是相对于什么参照系的?)因此发射后航天器的最小总能量近似为零。由于发射前航天器的总能量为 $-GMm/R$,因此需要由火箭提供给航天器的能量为 GMm/R ,相当于把航天器加速到 $v = (2GM/R)^{1/2} \approx 11.2$ 千米/秒的速度。这个速度被称为“第二宇宙速度”(second cosmic velocity),有时也被称为摆脱地球引力束缚所需的速度,它也是一个等价速度。

更进一步,倘若我们想把航天器发射得更远些,比方说发射到太阳系之外——就像本系列序言中所提到的“先驱者号”(Pioneer)探测器一样,火箭需要达到的速度又是多少呢?这个问题比前两个问题要复杂一些,因为所涉及的因素有地球与太阳两个星球的引力场,以及地球本身的运动。从太阳引力场的角度看,这个问题所问的其实就是在地球轨道所在处、相对于太阳的“第二宇宙速度”,即 $v = (2GM_s/R_{SE})^{1/2}$ (其中 M_s 为太阳质量, R_{SE} 为地球轨道的半径,也即太阳与地球之间的距离)^[8]。这一速度大约为42.1千米/秒。相对与第一、第二宇宙速度来说,这是一个很大的速度。但幸运的是,我们的地球本身就是一艘巨大的“宇宙飞船”,它环绕太阳飞行的速度约为29.8千米/秒。因此,如果航天器是沿着地球轨道运动的方向发射的,那么在远离地球时它相对于地球只要有 $v' = 42.1 - 29.8 = 12.3$ 千米/秒的速度就行了。在地心参照系中,发射这样的一个航天器所需要的能量为 $mv'^2/2 + GMm/R$ (其中后一项为克服地球引力场所需要的能量,即前面计算过的把航天器加速到第二宇宙

速度所需要的能量），相当于把航天器加速到 $v \approx 16.7$ 千米/秒的速度。这一速度被称为“第三宇宙速度”（third cosmic velocity），有时也被称为摆脱太阳引力束缚所需的速度，它同样也是一个等价速度，而且还是针对在地球上沿地球轨道运动方向发射航天器这一特殊情形的。

以上三个“宇宙速度”就是迄今为止火箭技术所跨越的三个阶梯。在关于“第三宇宙速度”的讨论中我们看到，行星本身的轨道运动速度对于把航天器发射到遥远的行星际及恒星际空间是很有帮助的。这种帮助不仅在发射时可以大大减少发射所需的能量，而且对于飞行中的航天器来说，倘若巧妙地安排航线，也可以起到“借力飞行”的作用，比如“旅行者号”就曾利用木星的引力场及轨道运动速度来进行加速。

三、齐奥尔科夫斯基公式

在上节中我们讨论了为发射不同类型的航天器，火箭所要达到的速度。与火箭之前的各种技术相比，这种速度是很高的。在早期的科幻小说中，人们曾设想用所谓的“超级大炮”来发射载人航天器。其中最著名的是法国科幻小说家凡尔纳（Jules Verne, 1828—1905年）的作品。凡尔纳在1865年发表的小说《从地球到月球》（*From the Earth to the Moon*）中曾经让三位宇航员挤在一枚与“神舟号”飞船的轨道舱差不多大的特制炮弹中，用一门炮管长达900英尺（约300米）的超级大炮发射到月球上去（最终没能击中月球，而成为了环绕月球运动的卫星）。不过，凡尔纳虽有非凡的想象力，却似乎缺乏必要的物理学及生理学知识。他所设想的超级大炮若真的在300米的炮管内把“炮弹”加速到11.2千米/秒（第二宇宙速度），则“炮弹”的平均加速度必须达到200 000米/秒²以上，也就是20 000g（ $g \approx 9.8 \text{米/秒}^2$ 为地球表面的引力加速度）以上。但是脆弱的人类身体所能承受的最大加速度只有不到10g。这两者之间的巨大差异无疑是灾难性的，因此凡尔纳的炮弹虽然制作精致，乘坐起来却一点也不会舒适。不仅不会舒适，且有性命之虞。事实上，英勇的宇航员们在“炮弹”出膛时早就变成了肉饼，炮弹最后有没有击中月球对他们都已不再重要了。而且若炮弹真的击中月球的话，其着陆方式属于所谓的“硬着陆”，就像陨石撞击地球一样，着陆时的速度差不多就是月球上的第二宇宙速度（约为2.4千米/秒），相当于在地球上从比珠穆朗玛峰还高30倍的山峰上摔到地面，这无疑是要把肉饼进一步摔成肉酱。

因此对于发射航天器（尤其是载人航天器）来说，很重要的一点就是航天器的加速过程必须发生在一个较长的时间里（减速过程也一样）。但是加速过程持续的时间越长，在加速过程中航天器所飞行的距

离也就越大。以凡尔纳的超级大炮为例，倘若炮弹的加速度小于 $10g$ ，则加速过程必须持续100秒以上，在这段时间内炮弹飞行的距离在500千米以上。炮弹的加速度越小，这段距离就越大。由于炮弹本身没有动力，因此这段距离必须都在炮管内。这就是说，凡尔纳超级大炮的炮管起码要有500千米长！建造这样规模的大炮显然是很困难的，别说凡尔纳时代的技术无法办到，即使在今天也是申请不到经费的。因此航天器的发射必须另辟蹊径^[9]。火箭便是一种与凡尔纳大炮完全不同但却非常有效的技术手段。

火箭是一种利用反冲作用推进的飞行器，即通过向与飞行相反的方向喷射物质而前进的飞行器。从物理学上讲这种飞行器所利用的是动量守恒定律。下面我们就来对火箭的飞行动力学作一个简单分析。

假设火箭在单位时间内喷射的物质质量为 $-dm/dt$ （ m 为火箭质量， $dm/dt < 0$ ），喷射物相对于火箭的速度大小为 u （方向与火箭飞行方向相反），则在时间间隔 dt 内，火箭的速度会因为喷射而得到一个增量 dv 。依据动量守恒定律，在火箭参照系中可以得到

$$mdv = -u dm$$

对上式积分并注意到火箭的初速度为零，便可得到

$$v = u \ln (m_i / m_f)$$

其中 m_i 与 m_f 分别为火箭的初始质量及推进过程完成后的质量（显然 $m_i > m_f$ ）。这一公式被称为齐奥尔科夫斯基公式（Tsiolkovsky formula），它是由上文提到过的俄国科学家齐奥尔科夫斯基发现的，时间是1897年，那时候的天空还是人类的“禁地”，连飞机都还没有上天

[10]。齐奥尔科夫斯基因为在航天领域的一系列卓越的开创性工作，而被许多人尊称为“航天之父”（father of astronautics）或“火箭之父”（father of rocketry）。

从齐奥尔科夫斯基公式中我们可以看到，火箭所能达到的速度可以远远地高于喷射物的喷射速度。这一点是很重要的，因为这意味着我们可以通过一种较低的喷射速度来达到航天器所需要的高速度，这在技术上远比直接达到高速度容易得多。从某种意义上讲，凡尔纳的超级大炮之所以没能成为一种载人航天器的发射装置，正是因为它试图直接达到航天器所需要的高速度。

但是火箭虽然能够达到远比喷射物喷射速度更高的速度，为此而付出的代价却也不小，因为火箭所要达到的速度越高，它的有效载荷就必须越小。这一点从齐奥尔科夫斯基公式中可以很容易地看到。我们可以把公式改写为 $m_f = m_i \exp(-v/u)$ ，由此可见，火箭的飞行速度 v 越高，它的有效载荷（ m_f 中的一部分）也就越小。假如我们想用 $u=1$ 千米/秒的喷射速度来达到第一宇宙速度（即将有效载荷送入近地轨道），则 $m_f/m_i \approx 0.00037$ ，也就是说一枚发射质量为1000吨的火箭只能让几百千克的有效载荷达到第一宇宙速度，这样的效率显然是太低下了。

为了克服这一困难，齐奥尔科夫斯基提出了多级火箭的设想。多级火箭的好处是在每一级火箭的燃料用尽后可以把该级火箭的外壳抛弃掉，从而减轻下一级火箭所负载的质量。在理论上，火箭的级数越多，运载效率就越高，不过在实际上，超过三级的火箭其技术复杂性的增加超过了运载效率上的优势，使用起来得不偿失。因此，目前我们使用的火箭大都是三级火箭。即便使用多级火箭，航天飞行的消耗依然是惊人的，通常一枚发射质量为几百吨的火箭只能将几吨的有效载荷送入近地

轨道，比如发射“神舟号”飞船的长征二号F型火箭发射质量约为480吨，近地轨道的有效载荷约为8吨。

四、接近光速

前面说过，这个星际旅行系列主要是为了讨论未来的星际旅行技术而写的，因此，在这里我们也要把目光放远些，看看上节讨论的火箭动力学在火箭速度持续提高，乃至接近光速时会如何。截至2013年7月，人类发射的航天器中飞得最远的是1977年9月5日发射的“旅行者一号”（Voyager 1）。经过近36年的漫长飞行，它已经飞到了离太阳约187亿千米处，远远超出了太阳系已知最外围的行星——海王星，或曾经最外围的行星——冥王星——的轨道。但是，这个距离跟离太阳最近的恒星——半人马座比邻星（Proxima Centauri）——的距离相比，还不到万分之五。由此可见，人类要想走得更远，必须要有更快的航天器。在齐奥尔科夫斯基公式中火箭的速度是没有上限的，通过提高喷射物的喷射速度，通过增加火箭质量中喷射物所占的比例，火箭在原则上可以达到任意高的速度。但是，这一点显然是错误的，因为物体的运动速度不可能超过光速，这是相对论的要求^[11]。这表明，当火箭的运动速度接近光速时，齐奥尔科夫斯基公式将不再成立。那么，有没有一个比齐奥尔科夫斯基公式更普遍的公式，在火箭运动速度接近光速时仍成立呢？这就是本节所要讨论的问题。

首先，简单的答案是：这样的公式是存在的。事实上，这样的公式不仅存在，而且并不复杂，因此我们干脆在这里把它推导出来，以满足大家的好奇心。这一推导所依据的基本原理仍然是动量守恒定律，我们也仍然在火箭参照系中计算火箭速度的增量。这里要补充说明的是，所谓火箭参照系，指的是所考虑的瞬间与火箭具有同样运动速度的惯性参照系（因此在不同的时刻，火箭参照系是不同的）。我们用带撇的符号表示火箭参照系中的物理量（这是讨论相对论问题的惯例）。与上节的

讨论相仿，假设火箭在单位时间内喷射的物质质量为 $-dm'/dt'$ （ m' 为火箭质量， $dm'/dt'<0$ ），喷射物相对于火箭的速度大小为 u （方向与火箭飞行方向相反），则在时间间隔 dt' 内，火箭的速度会因为喷射而得到一个增量 dv' 。依据动量守恒定律，在火箭参照系中可以得到

$$m'dv' = -u dm'$$

这里 dm' 为喷射物的相对论质量（运动质量），这一公式对于 u 接近甚至等于光速的情形也成立^[12]。在非相对论的情形下，上面所有带撇的物理量都等于静止参照系（地心参照系）中的物理量，因此对上述公式可以直接积分，这种积分的含义是对上式中的速度增量进行累加。但在相对论中，速度合成的规律是非线性的，把这些在不同时刻——因而在不同参照系中——的速度增量直接累加是没有意义的，因此上述速度增量必须先换算到静止参照系中才能积分。

运用相对论的速度合成公式， dv' 所对应的静止系中的速度增量为

$$dv = \frac{dv' + v}{1 + \frac{v dv'}{c^2}} - v = \left(1 - \frac{v^2}{c^2}\right) dv'$$

将这一结果与在火箭参照系中所得的关于 dv' 的公式联立可得

$$\frac{dv}{1 - \frac{v^2}{c^2}} = -u \frac{dm'}{m'}$$

对这一公式积分，并进行简单处理，便可得到

$$v = c \tanh\left(\frac{u}{c} \ln \frac{m_i}{m_f}\right)$$

其中火箭的初始质量 m_i 与推进过程完成后的质量 m_f 都是在火箭参照系中测量的。这就是齐奥尔科夫斯基公式在相对论条件下的推广。对于低速运动的火箭， $(u/c) \ln(m_i/m_f) \ll 1$ ，因而 $\tanh[(u/c) \ln(m_i/m_f)] \approx (u/c) \ln(m_i/m_f)$ ，上述公式退化为普通的齐奥尔科夫斯基公式。由于对于任意 x ， $\tanh x < 1$ ，因此由上述公式给出的速度在任何情况下都不会超过光速，从而符合相对论的要求。

上述公式的一个特例是 $u=c$ 的情形，即喷射物为光子（或其他无质量粒子）的情形。这种火箭常常出现在科幻小说中，通常是以物质与反物质的湮灭作为动力来源。对于这种情形，上述公式简化为：。如果将火箭90%的质量转化为能量作为动力，火箭的飞行速度可以达到光速的99%。

五、飞向深空

宇宙的浩瀚是星际旅行家们所面临的最基本的事实。即使能够达到接近光速的速度，飞越恒星际空间所需的时间仍然是极其漫长的。比如从太阳系出发，到银河系中心大约要3万年，到仙女座星云

（Andromeda Galaxy，也称为M31，为河外星系）大约要220万年，到室女座星系团（Virgo，为河外星系团）大约要6 000万年.....相对于人类弹指一瞬的短暂生命来说这些时间显然是太漫长了。但是且慢悲观，因为我们还有一个因素可以依赖，那就是相对论的时钟延缓效应。在相对论中运动参照系中的时间是由所谓的“本征时间”来表示的，它与静止参照系中的时间之间的关系为

$$\tau = \int \left(1 - \frac{v^2}{c^2}\right)^{1/2} dt$$

把这个公式运用到火箭参照系中， τ 就是宇航员所感受到的时间流逝。很显然，火箭的速度越接近光速，宇航员所感受到的时间流逝也就越缓慢。考虑到这个因素，宇航员是不是有可能在自己的有生之年到银河系中心、仙女座星云、甚至室女座星系团去旅行呢？下面我们就来计算一下。

我们考虑一个非常简单的情形，即火箭始终处于匀加速过程中。当然这个匀加速度是在火箭参照系中测量的。为了让宇航员有“宾至如归”的感觉，我们把加速度选为与地球表面的重力加速度一样，即 g 。用数学语言表示：

$$\frac{d^2 x'}{dt'^2} = g$$

把这一加速度变换到静止参照系（地心参照系）中可得

$$\frac{d^2 x}{dt^2} = \left(1 - \frac{v^2}{c^2}\right)^{3/2} g$$

由此积分可得

$$x = \frac{c^2}{g} \left[\left(1 + \frac{g^2 t^2}{c^2}\right)^{1/2} - 1 \right]$$

只要加速的时间足够长（即 $gt \gg c$ ），上式可近似为 $x \approx ct$ 。这表明在地心参照系中，经过长时间加速后飞船基本上是以光速飞行的。但是我们感兴趣的是宇航员所经历的时间，即“本征时间” τ ，这是很容易利用上式—— τ 的定义——计算出的，结果为（请读者自行验证）

$$\tau = \frac{c}{g} \operatorname{arcsinh} \left(\frac{gt}{c} \right)$$

我们可以从 τ 和 x 的表达式中消去 t ，由此得到

$$\tau = \frac{c}{g} \operatorname{arcsinh} \left\{ \left[\left(1 + \frac{gx}{c^2}\right)^2 - 1 \right]^{1/2} \right\}$$

如果 $x \ll c^2/g$ （约1光年），即飞行距离远小于1光年，上式可近似为： $\tau \approx (2x/g)^{1/2}$ ，这正是我们熟悉的非相对论匀加速运动的公式。如果 $x \gg c^2/g$ ，即飞行距离远大于1光年，上式可以近似为 $\tau \approx (c/g) \ln(2gx/c^2)$ 。下面我们将只考虑这种情形。考虑到抵达一个目的地

后，通常还要做一些考察研究、拍照留念的事情，因此火箭不能一味加速，而必须在航程的后半段进行减速，从而旅行所需的时间应当修正为（最右侧表达式中 τ 以年为单位， x 以光年为单位）

$$\tau = \frac{2c}{g} \ln\left(\frac{gx}{c^2}\right) \sim 2 \ln x$$

由这一公式不难看到：倘若旅行的目的地是银河系的中心， $x=30\ 000$ 光年，则 $\tau \sim 20$ 年。这就是说，在宇航员看来，仅仅20年的时间，他就可以到达银河系的中心，即使考虑到返航的时间，前后也只需40年的时间，他就可以衣锦还乡了。这就是相对论的奇妙结论！只不过，当他回到地球时，地球上的日历已经翻过了整整6万年，他的孙子的孙子的孙子……（如果有的话）都早已长眠于地下了^[13]。

运用同一公式，我们还可以计算出到达仙女座星云所需的时间约为29年，到达室女座星系团所需的时间约为36年……（在这里，读者们对于对数函数的增长之缓慢大概会有一个深刻印象吧。）倘若一个宇航员20岁时坐上火箭出发，如果他可以活到80岁，那么在他有生之年（不考虑返航——“壮士一去兮不复返”），他可以到达10 000 000 000 000（10万亿）光年远的地方。这个距离已经远远远远地超过了可观测宇宙的线度。因此，这样一位宇航员在其有生之年可以到达宇宙中任意远的地方！

由此看来，星际旅行似乎并不像人们渲染的那样困难。倘如此，则我们也就不必费心讨论什么虫洞（wormhole）和生命传输机（transporter）了，直接坐上火箭遨游太空就是了。事情当然并不如此简单，别忘了在我们的计算中火箭是一直在加速的（否则的话，那个帮了我们大忙的对数函数就会消失），那样的火箭所耗费的能量是惊人的

（究竟要耗费多少能量呢？运用本文给出的结果，读者可以自己试着计算一下）^[14]。不过这种能量耗费所带来的困难比起建造虫洞所面临的困难来终究还是要小得多。因此，运用那样的火箭探索深空也许真的会成为未来星际旅行家们的选择。唯一的遗憾是，他们只要走得稍远一点，我们就没法分享他们的旅行见闻了。

因为相对论只保佑他们，不保佑我们。

2003年10月14日写于纽约

2013年7月13日最新修订

[1]本文发表之后数小时，北京时间2003年10月15日早晨9时整，“神舟五号”飞船载着宇航员杨利伟从酒泉卫星发射中心发射升空。飞船升空587秒后与火箭分离，进入轨道倾角为42.4度、近地点高度为200千米、远地点高度为350千米的预定椭圆轨道。飞船飞行至第五圈时变轨进入高度为343千米的近地圆轨道。北京时间2003年10月16日早晨6时23分，飞船在环绕地球14圈后在内蒙古四子王旗北部的主着陆场安全着陆，不久杨利伟自主出舱。至此，我国第一次载人航天飞行取得圆满成功。杨利伟成为我国第一位进入太空的宇航员，我国成为继苏联与美国后第三个独立掌握载人航天技术的国家。“神舟五号”的发射是人类历史上的第241次载人航天飞行。杨利伟是人类历史上进入太空的第952人次。

[2]大气层与行星际空间是连续衔接的，所谓“穿过大气层”指的是穿过厚度在百余千米以内的相对稠密的大气层。

[3]当然，这里我们要忽略空气阻力，并且还要忽略地球表面的地形起伏。

[4]这里“卫星”指的是环绕地球运动的物体，其轨迹局限在有限区域内（否则的话，可能的轨迹将包括抛物线与双曲线）。同时我们还假定地球的引力场是一个严格的平方反比中心力场，且

忽略任何其他星体的引力场。

[5]确切地讲是指速度的大小，下文提到的“向心力”、“引力”等也往往指的是大小，请读者依据上下文自行判断。

[6]这里参照系的原点取在地心，且忽略了由地球自转导致的卫星动能（因此而带来的误差小于1%）。

[7]确切地讲是忽略太阳引力场中引力势能的变化。在这一限制之下其他行星的引力场也同样可以忽略。

[8]这里我们忽略了地球轨道的微小椭率，而将之视为圆轨道。

[9]类似于凡尔纳大炮那样的装置在表面引力较弱的星球——比如月球——上建造起来就会容易许多，因此曾有人设想它可以成为未来月球基地的航天器发射装置。

[10]这一公式的正式发表是在1903年，与莱特兄弟（Wright brothers）的飞机同一年。另外，新近发现的一些史料表明，英国皇家军事学院（Royal Military Academy）的科学家早在1813年就得到过类似的结果。

[11]在理论与实验上都有迹象表明，在特定的条件及特定的含义下，运动速度超过光速并非绝对不可能，但这种超光速并不像许多科普爱好者所认为的那样，是推翻了相对论。

[12]假如 u 等于光速，则 dm' 理解为 dE'/c^2 （ E' 为喷射物的能量）。

[13]这类结果早年曾引起过争议，并被称为“时钟佯谬”（clock paradox），但其实并无佯谬可言，感兴趣的读者可参阅拙作“关于时钟佯谬”（已收录于本书）。

[14]需要提醒读者的是，这种速度极其接近光速的火箭将会遇到的一个我们未曾提及的问题，那就是：它所经过的星际空间中的所有物质——哪怕细微到基本粒子——相对于火箭都具有极高的能量，从而有可能造成极大的危害。

生命传输机^[1]

看过科幻电视连续剧《星际迷航》（*Star Trek*）的人可能对剧中的生命传输机（Transporter）留有深刻的印象。需要进入别的飞船或在星球上着陆的飞船乘员站在生命传输机的控制室中，随着操作人员的一句“Energize”的口令，乘员的身体渐渐分解成了一片闪烁的粒子，从控制室中悄然消失；几乎与此同时，在传输目的地，一个粒子团魔术般地出现，并渐渐变得明亮起来，最终完整地复现出了飞船乘员（图14）。整个分解和复合的过程只需几秒钟。据说《星际迷航》的编导们最初设计这么一个生命传输机是为了省钱，因为当时摄制组的经费负担不起拍摄星际飞船在星球表面着陆所需的特技过程。



图14 生命传输机

像生命传输机那样的概念使许多人都感到了兴趣。念中学时我曾翻过一本由美国学者霍夫施塔特（Douglas R. Hofstadter）和丹尼特（Daniel C. Dennett）撰写的名为《心我论》（*The Mind's I*）的书，一开头就提到了类似于生命传输机的装置，由此展开了许多生命哲学方面的讨论。对研究星际旅行的人来说，像生命传输机那样的装置是让脆弱而短暂的生命以基本粒子的形式跨越星际间严酷的环境和近乎无限的时空尺度的理想手段。

《星际迷航》播映之后还出版了一本《技术手册》（*Technical Manual*），替剧中用到的许多新技术和新概念作了书面描述。从《技术手册》上看，《星际迷航》中的生命传输机是直接将组成原生命体的基本粒子传输到目的地进行复现的。按照我们对微观世界的了解，这是不必要的。因为依据量子力学的基本原理，同一类型的基本粒子彼此间是完全相同的。因此在使用生命传输机的过程中，组成生命体的那些基本粒子本身是否直接被传输到目的地其实并不重要，因为那些基本粒子本身并没有任何特殊性。真正需要传输的只是有关生命微观组成的完整信息^[2]。只要有了这些信息，通过什么途径，从什么地方获取复现生命体所需的基本粒子是无关紧要的。事实上，生命虽然奥妙，但组成生命体的那些基本粒子——注意不是分子，而是基本粒子——本身据我们所知在宇宙间是普遍存在的。因此，如果有一天星际旅行家们真的建造出了像生命传输机那样的装置，我们所要做的将只是设法把接收和复现装置送到目的地（《星际迷航》中连这些装置也省略了，看来经费的确是比较紧张），此后两地之间的旅行在原则上就可以像今天人们所熟悉的电波通信那样快捷和“方便”了。

那么像生命传输机那样能够把生命分解为基本粒子，并在异地完整

复现的装置在物理上是否可以实现呢？如果可以实现，它的作用过程是否会像人们在《星际迷航》中所看到的那样呢？这些就是本文所要讨论的问题。至于生命传输机所引发的有关生命哲学方面的思考则不在本文的考虑之列，感兴趣的朋友可以去看看《心我论》或其他类似的书。

按照前面的介绍，生命传输机在物理上能否实现的一个关键的环节，就在于能否获得有关生命微观结构的完整信息。我们不妨回想一下，在宏观世界里如果我们要复制一样东西，比方说一件家具，该怎么做？通常我们会从各个角度对所复制的家具进行观察，研究它的材料，分析它各部件的拼合方式，如此等等。从物理学的角度讲，所有这些都是对被复制的物体进行观测，复制过程所需的信息就来源于这些观测。这些观测所需达到的细微程度则显然与复制本身所需达到的精密程度密切相关。对于家具而言，人们关心的是它的外观、手感、强度等性质，复制物只要在这些性质上做到与原件难以区分就可以了。由于这些性质都是宏观性质，有关它们的信息都是宏观信息，因此为复制家具所需的观测是宏观意义上的观测，这样的观测在物理学上是没有任何原则性困难的。

那么复制生命的情况又如何呢？这里所说的复制生命不是今天大家正在热议的克隆（clone），克隆所复制的只是生命的躯壳，而我们讨论的是真正地、全息意义上的生命复制。这种复制不仅包括躯壳，还必须包括记忆、意识、情感、智慧等原生命体所具有的全部重要特征。这里我们遇到的第一个巨大的困难就是我们并不清楚生命——尤其是像人类这样的“高等”生命——的全部奥秘，比方说我们迄今还不了解意识的物理起源。我们不清楚人的意识以及其他许多深层功能的存在究竟是依赖于人体在哪个物质层次上的结构，是原子、分子层次？还是细胞层次？亦或干脆就是一种独立的存在？依据答案的不同，为传输生命所需获得

的有关生命结构的信息，以及在传输和复现生命过程中所需使用的物质基元（building block）将会有所不同。

很明显，在没有找到这些问题的真正答案之前是无法对复制生命的可行性做出准确判断的。不过从星际旅行的角度讲，如果生命传输机所需传输的是细胞（或细胞以上的组织），那么由于细胞本身就是一种初等的生命，在星际间的环境和时间跨度上维持它们与直接让人进行星际旅行所面临的困难也许只有程度上的差别，从而生命传输机对于星际旅行的价值就要大打折扣。本文将不讨论这种类型的生命传输机（《星际迷航》中的生命传输机显然也不是这一类型的）。另一方面，如果复制生命需要涉及非物质的东西（比方说如果意识是物质以外的独立存在），那么我们目前显然尚不具备讨论这一问题的物理学依据。

因此本文所要——或者说所能够——讨论的只有一种情形：即对生命的复制是在原子、分子或其他基本粒子层次上进行的。这也是生命传输机对星际旅行来说具有最大价值的情形（《星际迷航》中的生命传输机就属于这一类型）。因为正如前面所说，同一类型的基本粒子（或简单的粒子组合如原子、分子）在量子力学意义上是全同的，而且在这一层次上物质的组元（质子、电子等）在宇宙中是普遍存在的，这就使得直接传输组成生命的物质（以及维持这种物质）成为不必要，从而大大简化了生命传输机的结构。对于这种类型的生命传输机，只要我们能获得有关生命微观结构的完整信息，它的制造以及它在星际旅行中的使用至少在理论上就具有了相当大的可能性。

因此问题归结为我们是否有可能获得有关生命微观结构的完整信息。

在讨论如何获取有关生命微观结构的完整信息之前，让我们先来估

计一下这种信息的数量，以便大家有个概念。人体大约由一万亿亿亿（ 10^{28} ）个原子组成。假如对这一结构中每个原子的描述（包括它与周围原子的连接方式）平均需要100比特（byte）的信息，那么有关生命微观结构的完整信息大约有 10^{21} GB（一个GB约等于10亿比特）。

10^{21} GB的信息是个什么概念呢？打个比方吧，这样数量的信息，如果用容量为100GB的计算机硬盘来储存，大约需要1 000亿亿张硬盘。这些硬盘如果摆放起来的话，足以覆盖整个地球表面（不分陆地海洋）100遍！

传输和储存如此大量的数据本身无疑也是一个很大的挑战，但这种挑战相对于复制生命所面临的全部复杂性来说只不过是冰山之一角！

复制生命的真正复杂性来自这样一个事实：那就是获取一个体系微观上的完整信息在物理学上远不是一件轻而易举的事情，它和复制家具所涉及的获取体系的宏观信息有着本质的差别。这一差别来自于今年已逾百岁“高龄”的量子力学。一百多年前，自伽利略（Galileo Galilei）和牛顿（Isaac Newton）以来岿然屹立已达数百年之久的经典物理学大厦如同一串精巧的多米诺骨牌，被一朵“物理学晴朗天空中的小小乌云”——黑体辐射问题——撞了一下腰，竟尔轰然倒塌。所幸的是物理学本身就像浴火重生的火凤凰，从灰烬中脱胎出了一个崭新的领域，那便是量子力学。但是，对钟情于生命传输机的星际旅行家们来说，不幸的是：获取一个体系微观上的完整信息的美好愿望却被无情地压在了经典物理学的那片厚厚的废墟下面……

量子力学的出现导致了物理理论及其描述自然的总体方式的彻底变革。在量子力学中，对一个物理体系的描述由所谓的“波函数”（wave function）来表示^[3]。许多传统的经典物理学概念——比如粒子所在的

位置、粒子的运动速度，等等——失去了经典物理学赋予它们的实在性。量子力学诞生之后，尤其是著名的“不确定性原理”（uncertainty principle）提出前后，物理学家们对这一理论的内涵、它的自洽性和完备性等问题进行了长时间激烈的争论。那些争论大大澄清和加深了人们对许多量子力学基本概念的理解。从那些让物理学获益良多的争论中衍生出了许多全新的分支领域，其中的一个叫做量子力学测量理论，它是我们讨论获取一个体系微观上的完整信息的理论依据。

自测不准原理提出以来，物理学家们对量子力学测量理论的研究已经进行了整整四分之三个世纪。如果注意到这种研究是在量子力学的基本数学框架未出现重大变动的情况下进行的，并且有20世纪几乎所有最伟大的物理学家——比如爱因斯坦（Albert Einstein）、玻尔（Niels Bohr）、海森伯（Werner Heisenberg）、玻恩（Max Born）、薛定谔（Erwin Schrödinger）等——的积极参与，却直到今天也没能形成一个被普遍认可的理论，这在科学史上是颇为罕见的。量子力学在概念层次上的微妙性由此可见。量子力学的初学者们常常被告诫：“如果初学量子力学就觉得明白了，那你一定是没有理解它。”在量子力学炽热发展的时期，新的理论模型层出不穷。据说当时评判一个新理论是否正确的“标准”之一就是看这个理论是否足够“疯狂”，如果不是，那它一定是错的！全面地讨论量子力学测量理论远远超出了本文的范围。不过，值得庆幸的是虽然并不存在一个被普遍认可的测量理论，但分歧主要是集中在对理论的诠释上，物理学家们对测量理论的一些主要结论还是相当程度的共识的。简单地说，量子力学测量理论有别于经典测量理论的一个最基本的特点就是：观测过程本身对被观测体系造成的干扰是不可忽略的。用一句许多量子物理学家喜爱的俗语来表述就是：在量子力学这部大戏中，观测者既是观众也是演员。

量子测量理论的这一特点对获取有关生命微观结构的完整信息会造成一个很棘手的问题，那就是体系的微观状态经过一次测量就会发生变化。而状态一变，此后的测量所获得的就不再是关于体系原先微观状态的信息了。这就是说对一个体系的微观状态只能进行一次有效的测量。当然，“一次测量”在逻辑上并不意味着就只能得到“一点点”信息，我们也许可以期盼某种非常“聪明”的测量方法，一次就可以得到一个量子体系的全部信息。不幸的是，量子力学测量理论的另一个著名的结论就是：有一些可观测量是相互排斥，从而不可能在一次测量中同时获得精确结果的。换句话说，对一个量子体系的单次测量所能得到的信息往往注定只能是不完整的！

在研究普通的量子体系——比如氢原子——时这一点并不造成实质的困难，因为自然界中所有的氢原子都是一样的。我们可以对许多氢原子进行独立的测量，然后对结果进行综合分析。这正是对一个量子体系进行测量的标准方法。事实上在考虑量子力学测量问题时人们通常引进所谓的“系综”（ensemble）——即大量全同体系的集合——的概念，对一个量子力学体系的测量事实上是针对系综中各个全同体系进行大量的独立测量。这些独立测量的结果的统计分布由体系的波函数所描述^[4]。反过来，通过选择适当的待测物理量或物理量的组合，对一个系综中各个全同体系进行充分多的独立测量，从测量结果中原则上也可以反推出体系的波函数来。而波函数一旦确定，在量子力学意义上也就获得了有关体系微观结构的完整信息。

很明显，把这套理论用到我们所讨论的获得有关生命微观结构的完整信息的问题上来就会陷入一种“先有鸡还是先有蛋”的循环之中。因为按照上述理论，为了获取关于某个生命体微观结构的完整信息，必须先制备一个关于这一生命体的系综。但是生命体不像氢原子那样具有微观

全同性，自然界中根本就不存在关于生命体的系综。这就意味着要想制备一个关于生命体的系综，我们必须自行复制生命体。而为了能够复制一个生命体，我们就需要先知道关于该生命体微观结构的完整信息。

绕了一圈我们依然两手空空。

因此，获得有关生命微观结构的完整信息按照我们今天对量子力学规律的理解是不可能的。如果复制生命——从而制造生命传输机——果真严格依赖于有关生命微观结构的完整信息，那它就同样是不可能的。不过“幸运”的是，虽然我们并不清楚生命——包括记忆、意识、情感、智慧等全部内涵——对微观结构的确切依赖程度，但这种依赖必定带有某些程度的模糊性。也就是说微观状态的某些程度的改变不会影响生命的任何本质特征。比方说头上缺几根头发，皮肤上多一两点色斑，身上少几个细胞等所对应的微观状态的差异显然都不会妨碍所复制生命的有效性。因此我们所需回答的问题可以弱化为：考虑到所有可被允许的模糊性，是否有可能获得复制生命所必须的微观信息？遗憾的是，对这一问题我们目前只能用一个双重的“无可奉告”来回答。因为我们既不清楚“可被允许的模糊性”的确切含义，也没有对量子力学测量理论研究到足以回答这类问题的透彻程度。我们比较有把握的结论是：在简单意义上精确复制生命——即复制生命的全部微观结构——的生命传输机是不可能制造的。

最后我们再讨论一下如果生命传输机存在，它的工作情形是否会像图14所示的那样干净利落，在几秒钟之内点尘不惊地完成复制过程。当然，我们不可能讨论生命传输机的具体工作方式，我们只想来计算一下把一个人分解为基本粒子或由基本粒子复合成一个人所需吸收或释放的能量。假如生命传输机对人体的分解和复合是在亚原子——即质子、中子、电子等——的层次上进行的，那么人体将会被分解为大约10万亿亿

亿（ 10^{29} ）个亚原子粒子（比上文提到的原子数目多一个数量级左右）。由于平均每个亚原子粒子的结合能约为1兆电子伏特（1Mev），因此分解（复合）过程所需吸收（释放）的能量大约为1亿亿焦耳

（ 10^{16}J ），这相当于100万吨TNT炸药爆炸时释放的总能量！因此生命传输机操作人员的那句冷静而平淡的“Energize”背后所蕴含的能量其实是与核爆炸中令天地为之变色的蘑菇云所象征的能量不相上下。这种类型的生命传输机的作用过程——尤其是复合过程——是很难如电视上那样点尘不惊的。当然，如果生命传输机只是在原子或分子层次上对人体进行分解和复合，所涉及的能量就会小得多，大约相当于几十到几百公斤TNT炸药爆炸时释放的能量^[5]。一般来说，生命传输机对生命体的分解与复合所涉及物质层次越低，在分解与复合过程中吸收与释放的能量就越多。

我们关于生命传输机的讨论到这里就结束了，与星际旅行中的另一个流行的方案——虫洞——相比，生命传输机在理论可行性方面似乎略显乐观。但我们必须看到，这种乐观性在很大程度上是建立在对生命本质的无知之上的，就像在相对论之前人们可以乐观地认为运动速度在原则上是不受限制的。科学是美丽的，它受益于我们的想象力，又转而为想象力插上新的翅膀。但科学同时也是严谨的，它并不是漫无边际的想象。对生命本质的无知绝不是我们乐观的理由。如果我们真的想要寻求一点乐观的话，也许时间是最好的乐观理由，因为《星际迷航》的故事——确切地说是我所看过的那部分故事——发生在24世纪，我们还有300年的时间来更好地理解生命，理解物理学。也许到那时我们会更好地理解生命传输机——无论它是可行的还是不可行的。

2003年1月2日写于纽约

[1]本文曾发表于《科学画报》2003年第10期（上海科学技术出版社出版）。

[2]在后文中将会提到，对这里所说的“完整”两字不宜理解得过于绝对。

[3]确切地说，在量子力学中，对一个物理体系的描述体现在所谓的“状态”（state）上，“波函数”是状态在具体表象——比如坐标表象——下的函数表示。

[4]这里所说的系综理论只是量子力学测量理论所涉及的若干种诠释中的一种，但可以算是最直接对应于量子力学数学体系的诠释。

[5]对于爱思考的朋友来说，这一数值是不需要计算就可以得出的。因为普通TNT炸药利用的不是别的，正是爆炸物在原子和分子层次上的结合能（叫做化学能）。因此把人体在这一尺度上分解或复合所涉及的能量大致就等于与人体质量相当的TNT炸药所能释放的能量。

虫洞：遥远的天梯

一、引言

1985年的一个学期末，加州理工大学（California Institute of Technology）的理论物理学教授索恩（Kip S. Thorne）刚刚上完一学年的课，正慵懒地靠在办公室的椅子上休息，电话铃声忽然响了起来。打来电话的是他的老朋友，著名行星天文学家萨根（Carl Sagan）。萨根当时正在撰写一部描写人类与外星生命首次接触的科幻小说。写作已近尾声，但身为科学家的萨根希望自己的作品——虽然只是一部科幻小说——尽可能地不与已知的物理学理论相矛盾。在这部小说中，萨根安排女主人公通过黑洞（blackhole）穿越了26光年的距离，到达遥远的织女星（Vega）附近。这是整部小说中最具震撼力的情节，但从物理学的角度看，却也是最可疑的细节。于是萨根打电话给从事引力研究的索恩，为这一细节寻求技术咨询。在经过一番思考和粗略的计算后，索恩告诉萨根：黑洞是无法用做星际旅行的工具的。他建议萨根使用虫洞（wormhole）这一概念，这便有了随后出版，并被拍成电影的著名科幻小说《接触》（*Contact*）。

萨根的小说顺利地出版了，索恩对虫洞的思考却没有因此而结束。

三年后，索恩和他的学生莫里斯（Mike Morris）在《美国物理杂志》（*American Journal of Physics*）上发表了一篇题为《时空中的虫洞

及其在星际旅行中的用途》（*Wormhole in spacetime and their use for interstellar travel*）的论文，由此开创了对所谓可穿越虫洞（traversable wormhole）进行理论研究的先河^[1]。作为教学性刊物的《美国物理杂志》也因此有幸在一个全新研究领域的开创上留下了值得纪念的一笔。

莫里斯和索恩的文章在虫洞研究中具有奠基性的意义，不过虫洞这一概念却并非他们两人首先提出的。早在1957年，美国物理学家惠勒（John Archibald Wheeler）和学生米斯纳（Charles W. Misner）就在一篇文章中提出了这一概念。那篇文章讨论的主题是所谓的“几何动力学”（geometrodynamics），那是一种试图把物理学几何化的理论。米斯纳和惠勒的“几何动力学”后来并没有走得很远，但他们在文章中提出的虫洞这一概念却在事隔30多年后得到了全新的发展，并成为了以星际旅行为题材的科幻小说的标准词汇，可谓是“有心栽花花不开，无心插柳柳成荫”。

二、什么是虫洞？

那么究竟什么是虫洞呢？形象地说，虫洞是连接两个空间区域的一种“柄”状的结构。图15便是一种很流行的虫洞图示，图中倒U字形曲面代表我们生活在其中的空间，连接两个空间区域A和B的直线段代表的便是这种“柄”状结构，即虫洞。图15是一种抽象化的图示，连接A和B的直线段实际上代表的是具有一定线度的结构。不难看到，由于

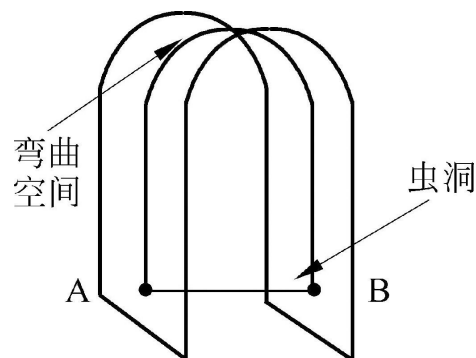


图15 一种典型的虫洞

这种“柄”状结构的存在，在A和B之间存在着两种不同类型的路径：一种由曲线表示，代表在普通空间中的路径；另一种由直线段表示，代表由于虫洞的存在而形成的新路径。由图15可以看到，沿直线段从A到B显然要比沿曲线近得多。通常科幻小说——包括前面提到的萨根的小说《接触》——所描述的通过虫洞的星际旅行，就是沿图中直线段进行的。

在虫洞的研究中，图15所示的虫洞被称为“宇宙内虫洞”（intra-universe wormhole），它连接的是同一个宇宙中两个不同的空间区域。除此之外，在理论上还有一类所谓的“宇宙间虫洞”（inter-universe wormhole），所连接的是两个不同的宇宙。科幻小说中的虫洞通常属于前一类。不过由于这两类虫洞的差别仅在于空间的大范围拓扑结构，对于讨论虫洞本身的结构来说，它属于哪一类并不重要。

在进一步讨论虫洞之前，我们先来澄清一个或多或少存在于文献中

的概念误区（或者说即便在文献作者的心中并无误区，却特别容易在读者之中造成误会的概念），那就是虫洞的存在并不意味着它们就一定是空间中的捷径（short-cut）。换句话说，虫洞的存在并

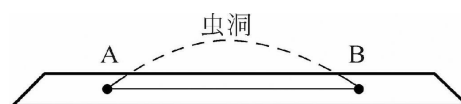


图16 另一种虫洞

并不意味着它们就一定能提供一种有意义的星际旅行路径。仔细观察图15不难发现，虫洞之所以成为连接A和B之间的捷径，完全是由于空间弯曲成了倒U字形所致。按照广义相对论，空间（确切地说是时空）的弯曲是由物质分布决定的，因而图15所表示的虫洞除了虫洞本身外，还对远离虫洞的背景空间中的物质分布作了十分苛刻的假定。如果不做这种相当人为的苛刻假定，虫洞的结构更有可能类似于图16所示。在图16中，由虫洞所形成的连接A和B的路径（即虚线路径）要比普通空间中的路径更长。很明显，利用图16所示的虫洞进行A和B之间的星际旅行是很不明智的。因此在概念上，虫洞并不等同于星际旅行的捷径。

三、萨根式的问题

尽管如此，虫洞无论对于物理学家、天文学家，还是星际旅行家来说，都依然是一个极富魅力的概念。前面提到的行星天文学家萨根对星际旅行所涉及的许多问题有一种很独特的提法，即从一个无限发达的文明（infinitely advanced civilization）的角度来看待星际旅行问题的可行性。对于虫洞，一个“萨根式”的问题可以表述为：

一个无限发达的文明是否有可能利用虫洞作为星际旅行的工具？

萨根所谓的“无限发达的文明”指的是在物理规律许可的情况下拥有一切能力的智慧生命。对于这样的智慧生命来说，图15和图16所示的虫洞或许并无实质区别。只要虫洞存在，即便其结构如图16所示，他们或许也有能力通过改变背景空间的曲率使之变为图15的形式。因此在这种“萨根式”的问题中，背景空间的具体结构有可能并不重要。

要利用虫洞作为星际旅行的工具当然首先得要有虫洞。宇宙间究竟有没有虫洞呢？这归根结底是一个观测问题。但起码到目前为止的答案是令人失望的，那就是迄今并未发现任何有关虫洞存在的直接或间接证据。因此现阶段我们对虫洞的探讨仅限于理论范畴。自莫里斯和索恩以来，物理学家们在对虫洞的研究上又获得了一些重要结果。这些结果主要是在有关引力和时空的经典理论——广义相对论——的框架内获得的。经过近一个世纪的研究，物理学家们对广义相对论的数学结构已经了解得相当透彻。尤其是自20世纪60年代以来，随着现代微分几何手段的应用，许多非常普遍的命题被相继证明，其中的一些对于虫洞研究有着十分重要的意义。

为了获得可作为星际旅行工具的虫洞，一个无限发达的文明可作两方面的努力：

（1）如果宇宙中不存在虫洞，他们可以试图“创造”虫洞。

（2）如果宇宙中存在虫洞，他们可以试图“改造”虫洞，使之适合于星际旅行的需要。

下面我们就分头介绍一下这两方面的努力。

四、虫洞的“创世记”——恼人的因果律

先来谈谈第一方面的努力，即“创造”虫洞。

所谓“创造”虫洞，指的是在原本没有虫洞的空间区域中产生出虫洞来。我们已经知道，虫洞是空间中的一种“柄”状结构，在拓扑学上具有这种“柄”状结构的空间被称为是复连通的，没有“柄”状结构（即没有虫洞）的普通空间则是单连通的。因此从拓扑学的角度讲，“创造”虫洞意味着使空间的拓扑结构发生变化。

那么空间的拓扑结构有可能发生变化吗？物理学家们对此进行了一系列的研究。1992年，著名英国理论物理学家霍金（Stephen Hawking）证明了这样一个定理。

[定理] 在广义相对论中，如果空间的拓扑结构在一个有界的区域内发生了变化，那么在这个变化所发生的时空范围内存在闭合类时曲线。

不熟悉相对论的朋友可能不知道什么叫做“类时曲线”（timelike curve）。在相对论中，类时曲线是物理上可以实现的有质量物体在时空中的运动轨迹。一个物体在空间中的运动轨迹闭合是十分寻常的事情，比如钟摆的运动，行星的运动，其在空间中的运动轨迹在适当的参照系中都是（近似）闭合的。但一个物理上可以实现的运动在时空中的运动轨迹闭合（即形成所谓“闭合类时曲线”）却是非同小可的事情。因为时空中的轨迹不仅记录了运动所经过的所有空间位置，而且还记录了它经过各空间位置的时刻。因此时空轨迹的闭合意味着不仅在空间上回到原点，而且在时间上也回到原点。换句话说，时空轨迹的闭合意味着

时间失去了实际意义上的单向性，或者说构造时间机器成为了可能！

我们都知道，自然万物的演化具有明显的不可逆性，最直接的经验莫过于我们的生命本身，从出生到成长，从衰老到死亡，每一步都不可抗拒、无可逆转。时间的单向性是物理学乃至全部自然科学中最基本的观测事实之一。如果时间不是单向的，那么物理世界中的因果关系也将不复存在，因为一个逆时间而行的旅行者可以在“结果”发生之后返回过去将产生结果的“原因”破坏掉^[2]。

因此霍金所证明的定理可以通俗地表述为：

[定理（通俗版）] 在广义相对论中，“创造”虫洞意味着放弃因果律。

如果放弃因果律，那么不仅物理学的大部分将会被改写，连科学本身的存在都将受到挑战。因为科学本质上就源于人类对自然现象追根溯源的努力，而正是因果律的存在使得这种努力成为可能。因此，依据霍金所证明的上述定理，在有足够证据表明因果律可以被破坏之前，我们必须认为改变空间的拓扑结构（即“创造”虫洞）是被广义相对论所禁止的。

广义相对论是现代物理学中最优美的理论之一，是引力理论和现代时空观念的基石，但它只是一个经典理论。物理学家们普遍认为，对引力和时空的真正描述就像对宇宙中其他基本相互作用的描述一样，必须是量子化的。对广义相对论的量子化被称为量子引力理论。

那么在量子引力理论中情况又如何呢？

早在量子理论出现之初物理学家们就已发现，许多被经典理论所禁

止的过程在量子理论中会成为可能，比如电子有可能出现在经典理论不允许出现的区域中。由此带来的一个很自然的问题就是：空间拓扑结构的改变会有幸成为这种量子过程“大家庭”中的一员吗？遗憾的是，对这一问题目前还没有明确答案。引力的量子化是当今理论物理面临的最困难的问题之一，迄今为止不仅尚未建立完整的理论，连一些基本的出发点也还在争议之中。在对量子引力理论的早期研究中，人们曾经设想时空就像海面一样，从大尺度上看平滑如镜，随着尺度的缩小渐渐显出起伏，当尺度缩小到一定程度时，就可以看到汹涌的波涛和飞散的泡沫。这个极小的尺度被称为普朗克尺度（Planck scale）。按照这种设想，在普朗克尺度上时空的结构会出现剧烈的量子涨落，不仅空间的拓扑结构可以发生变化，甚至还会产生所谓的时空泡沫（spacetime foam）。

但是，这种有关量子时空的直观设想在量子引力理论的各个具体方案中均遇到了不同程度的困难。初步的分析表明，量子引力理论并不完全禁止空间拓扑结构的改变，但是由产生虫洞所导致的空间拓扑结构的改变即使在量子引力理论中也极有可能是被禁止的。

因此我们可以有保留地认为，就目前我们所了解的物理学规律而言，“创造”虫洞有可能是一件连无限发达的文明也无法做到的事情。

五、虫洞工程学——负能量的困惑

接下来谈谈第二方面的努力，即“改造”虫洞，使之适合于星际旅行的需要。

即便“创造”虫洞是不可能的，一个无限发达的文明仍然可以通过改造宇宙中已经存在的虫洞（如果有的话），使之成为可穿越虫洞^[3]。这并不改变空间的拓扑结构，从而不违背任何禁止空间拓扑结构改变的物理学定理。

那么，改造一个可穿越虫洞——或者更具现实意义地说，维持一个改造后的可穿越虫洞——需要什么样的条件呢？

前面提到的莫里斯和索恩的文章的主要贡献就是对这一问题进行了定量的分析。他们研究了维持一个稳定的球对称虫洞所需要的物质分布。所谓球对称虫洞，指的是虫洞的出入口——即俗称为“嘴巴”（mouth）的部位——是球对称的。莫里斯和索恩发现，为了维持这样一个虫洞，在虫洞所形成的通道的最窄处——即俗称为“喉咙”（throat）的部位——必须存在负能量的物质

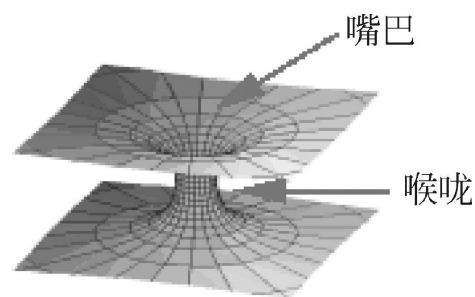


图17 虫洞的结构

（图17）。莫里斯和索恩的分析虽然对虫洞作了球对称这样一个简化假设，但是运用广义相对论及现代微分几何手段所做的进一步研究表明，他们得出的维持虫洞需要负能量物质的结论却是普遍成立的。

因此，想当一名虫洞工程师，首先得有负能量物质。

那么，什么是负能量物质呢？举一个简单的例子来说，学过牛顿定律的人都知道，用力推一个箱子，箱子就会沿推力的方向运动，推力的大小等于运动的加速度与箱子质量的乘积（假定阻力可以忽略）。这是大家熟悉的结果^[4]。但假如把箱子换成虫洞工程师的负能量箱子，情况就大不相同了。由于负能量箱子的质量小于零，若牛顿定律还能套用的话，加速度与推力的方向就变得彼此相反了。这表明你用力去推一个负能量箱子，非但不能把它推开，箱子反而会朝你滑过来！显然我们谁也没见过这么古怪的箱子，迄今为止人类在宏观世界中发现的所有物质都具有正能量，物质越多，通常能量也就越高。按照定义，只有一无所有的真空的能量才为零，而负能量意味着比一无所有的真空具有“更少”的物质，这在经典物理学中是近乎于自相矛盾的说法。

但量子理论的发展彻底改变了经典物理学关于真空的观念。在量子理论中，真空不仅具有极为复杂的结构，而且是高度动态的，每时每刻都有大量的虚粒子对产生和湮灭。在这种全新的真空图景下，负能量至少在概念层面上不再是不可思议的了。事实上，早在1948年，荷兰物理学家卡西米尔（Hendrik Casimir）就在理论研究中发现真空中两个平行导体板之间会出现负的能量密度，并由此预言了存在于这样一对导体板之间的一种微弱的相互作用。后来人们在实验上定量地证实了这种被称为卡西米尔效应（Casimir effect）的相互作用，从而间接地为负能量的存在提供了证据。20世纪70年代，霍金等物理学家在研究黑洞的辐射效应时发现，在黑洞的事件视界（event horizon）附近也会出现负的能量密度。20世纪80年代，物理学家们又发现了所谓的压缩真空（squeezed vacuum），即量子态分布异常的真空，在这种真空的某些区域中同样会出现负的能量密度。

所有这些令人兴奋的研究都表明，宇宙中看来的确是存在负能量物

质的。

但可惜的是，仅仅存在是不够的，还有数量的问题需要考虑。这方面的结果却极不容乐观，因为迄今所知的所有负能量物质都是由量子效应产生的，从而数量极其微小。拿卡西米尔效应来说，计算表明，一对平行导体板之间的负能量所对应的质量密度 ρ 大约为（其中 ρ 以千克每立方米为单位，平行导体板的间距 d 以米为单位）

$$\rho \approx -\frac{10^{-44}}{d^4}$$

这个结果表明如果平行导体板间距为一米的话，所产生的负能量的质量密度只有 10^{-44} 千克每立方米，相当于在每10亿亿立方米的体积内才有相当于一个基本粒子质量的负能量物质！

其他量子效应产生的负能量密度也大致相仿，只需把平行导体板间距换成那些效应所涉及的空间尺度即可。由于负能量的密度与空间尺度的四次方成反比，因此在任何宏观尺度上由量子效应产生的负能量都是微乎其微的。

另一方面，物理学家们对维持一个可穿越虫洞所需要的负能量物质的数量 M 也做了估算，结果发现（ M 以地球质量为单位，虫洞半径 R 以厘米为单位）：

$$M \approx -R$$

也就是说仅仅为了维持一个半径为一厘米的虫洞^[5]，就需要相当于整个地球质量的负能量物质！而且虫洞的半径越大，所需的负能量物质就越多。为了维持一个半径为一千米的虫洞所需的负能量物质的数量竟相当

于整个太阳系的质量！

这无疑是一个令所有虫洞工程师头疼的结果。因为一方面，迄今知道的所有产生负能量物质的效应都是量子效应，所产生的负能量物质的数量即使用微观尺度来衡量也是极其微小的。而另一方面，为了维持任何宏观意义上的虫洞所需的负能量物质的数量却是一个天文数字！

六、穿越虫洞——张力的挑战

虽然数字看起来不那么乐观，但是别忘了我们是在考虑一个“萨根式”的问题。我们的想象力已经无数次地低估过人类自身科学技术的发展，因此让我们姑且对来自“无限发达的文明”的虫洞工程师的技术水平做一个比较乐观的估计：假定他们利用某种远不为我们所知的技术手段真的获得了相当于整个太阳系质量的负能量物质，并成功地维持住了一个半径为1 000米的虫洞。

他们是否就可以利用这样的虫洞进行星际旅行了呢？

初看起来，半径1 000米的虫洞似乎应当满足星际旅行的要求了，因为1 000米的半径在几何尺度上已经足以让相当规模的星际飞船通过了。看过科幻电影的人可能对星际飞船穿越虫洞的特技处理留有深刻印象。从屏幕上，飞船穿越的似乎是时空中一条狭小的通道，飞船周围充斥着由来自遥远天际的星光和辐射组成的无限绚丽的视觉幻象（图18）。

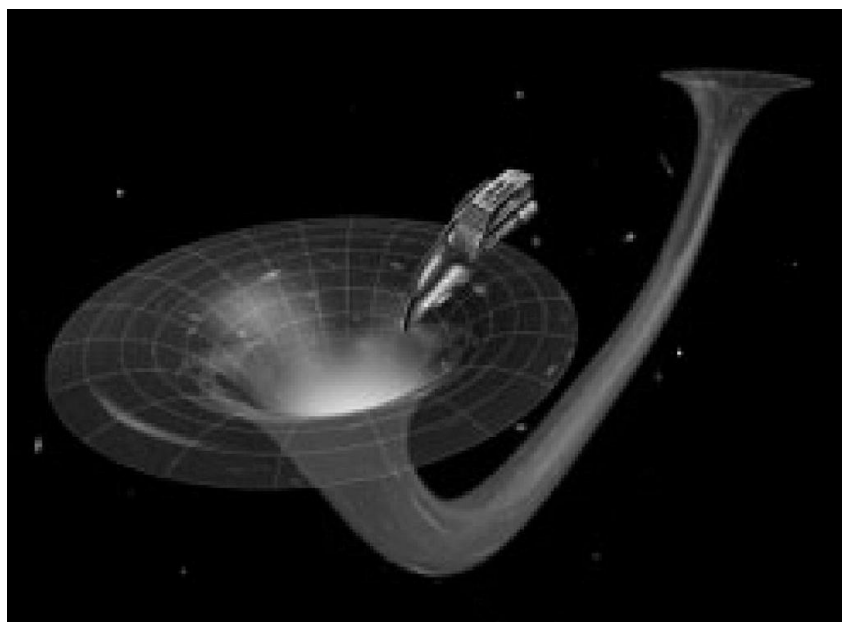


图18 星际飞船进入虫洞

但实际情况远没有那样诗情画意。

事实上，为了能让飞船及其乘员安全地穿越虫洞，几何半径的大小并不是星际旅行家所要考虑的主要问题。按照广义相对论，为了维持像虫洞那样时空高度弯曲的结构，必须依靠由负能量物质提供的巨大张力。而当飞船及其乘员穿越虫洞，尤其是穿越负能量物质密集的区域——即虫洞的“喉咙”部位——时，将几乎无可避免地会遭遇到这种张力。由于无论飞船还是飞船乘员，他们所能承受的张力都是有限的，因此穿越虫洞时所会遭遇到的张力大小对于星际旅行来说是至关重要的。那么这种张力究竟有多大呢？以球对称的虫洞为例，计算表明，在虫洞的“喉咙”部位，张力的大小约为

张力 \approx （物质所能承受的最大张力）/（以光年为单位的虫洞半径的平方）这里“物质所能承受的最大张力”指的是物质中的原子结构所能承受的最大张力。超越了这一极限，连组成物质的原子都将受到破坏，更遑论像飞船或飞船乘员那样的宏观物质了。这恐怕是任何程度的文明

——只要他们的生存还离不开物质形体——都很难突破的物理极限。从上述结果中我们看到，穿越虫洞所会遭遇到的张力大小与虫洞半径的平方成反比，虫洞的半径越大，张力就越小，从而也就越适合于作为星际旅行的通道。特别需要看到的是，半径小于一光年的球对称虫洞由于穿越时所会遭遇到的张力大小超过物质所能承受张力的理论极限，将很可能无法作为星际旅行的通道。

虽然以上都是比较粗略的估算，具体数值会因虫洞结构的不同而有所不同。但在数量级的意义上，这种估算已足以使我们看到维持一个可供星际旅行用的虫洞所面临的巨大的“工程学”困难，那就是：一方面，为了能让星际飞船安全通过，虫洞的半径至少要在1光年以上；另一方面，我们在前面已经介绍过，维持一个半径1千米的球对称虫洞所需的负能量物质数量约相当于整个太阳系的质量，且半径越大，所需的负能量物质也越多（与半径成正比），而1光年大约是10万亿千米，因此维持一个半径1光年的球对称虫洞所需的负能量物质数量约相当于太阳系质量的10万亿倍！

“太阳系质量的10万亿倍”是个什么概念呢？我们知道，整个银河系中所有发光星体的总质量大约是太阳系质量的1 000亿倍，因此维持一个可供星际旅行用的最小的球对称虫洞所需的负能量物质数量约相当于银河系中的所有发光星体质量总和的100倍！如果考虑到生物体所能承受的张力要远小于理论极限，对虫洞半径的要求将更高，所需的负能量物质的数量则将比上述估计值更大。使用数量如此惊人的物质，别说这些物质都是迄今尚未在任何宏观尺度上被发现的负能量物质，即便是普通的物质，也是近乎于天方夜谭的想法。

总体来说，目前还不清楚存在于微观尺度上的负能量物质是否有可能积累成宏观数量，如果这种积累是可能的，那么将一个已经存在的虫

洞改造并维持成适合星际旅行的虫洞在纯理论上是可能的。但改造并维持那样的虫洞所需的负能量物质的数量即便从宇宙学尺度上看也是极其惊人的。这种数量对于任何存在于我们这个宇宙中的文明——哪怕是无限发达的文明——来说，恐怕都是工程学上一个不可逾越的困难。

七、结语——遥远的天梯

在我们即将结束对虫洞的讨论时^[6]，我想起了远古神话中关于“天梯”（ladder to heaven）的一些传说。在远古的年代里，很多人幻想着天空中有一个圣洁而永恒的天堂，人的灵魂能在那里得到永生。虽然谁也不确定天堂离我们有多远，但有些人幻想着存在一些神秘的地方，人们可以从那里攀上天堂，那便是有关“天梯”的传说。古埃及的法老们曾经相信宏伟的金字塔可以成为他们的天梯；藏民们的一种传说，则认为天梯是神山上的一株巨树。从某种意义上讲，虫洞仿佛是一种现代版的“天梯”，一端连着古老而执着的梦想，一端连着遥远而璀璨的星空。

梦想与现实往往是有距离的，任凭虔诚的信徒们千百年不懈地期盼和寻觅，传说中的天梯终究没有被找到。人类对可穿越虫洞的研究才进行了短短十几个年头，下断语还为时过早。但从迄今所得的结果来看，利用虫洞进行星际旅行大致是介于“理论上不可能”和“实际上不可能”之间。在能够想象得到的将来，利用虫洞进行星际旅行很可能就像寻找遥远的天梯一样，只能是一个美丽却难圆的梦。

2002年9月26日写于纽约

2014年12月4日最新修订

[1]所谓“可穿越虫洞”，广义地讲，是指允许光信号穿越的虫洞；狭义地讲，则是指允许星际飞船穿越的虫洞。本文所讨论的是后一种。

[2]严格地讲，时间的非单向性（或闭合类时曲线的出现）并不一定导致因果律的破坏。有些物

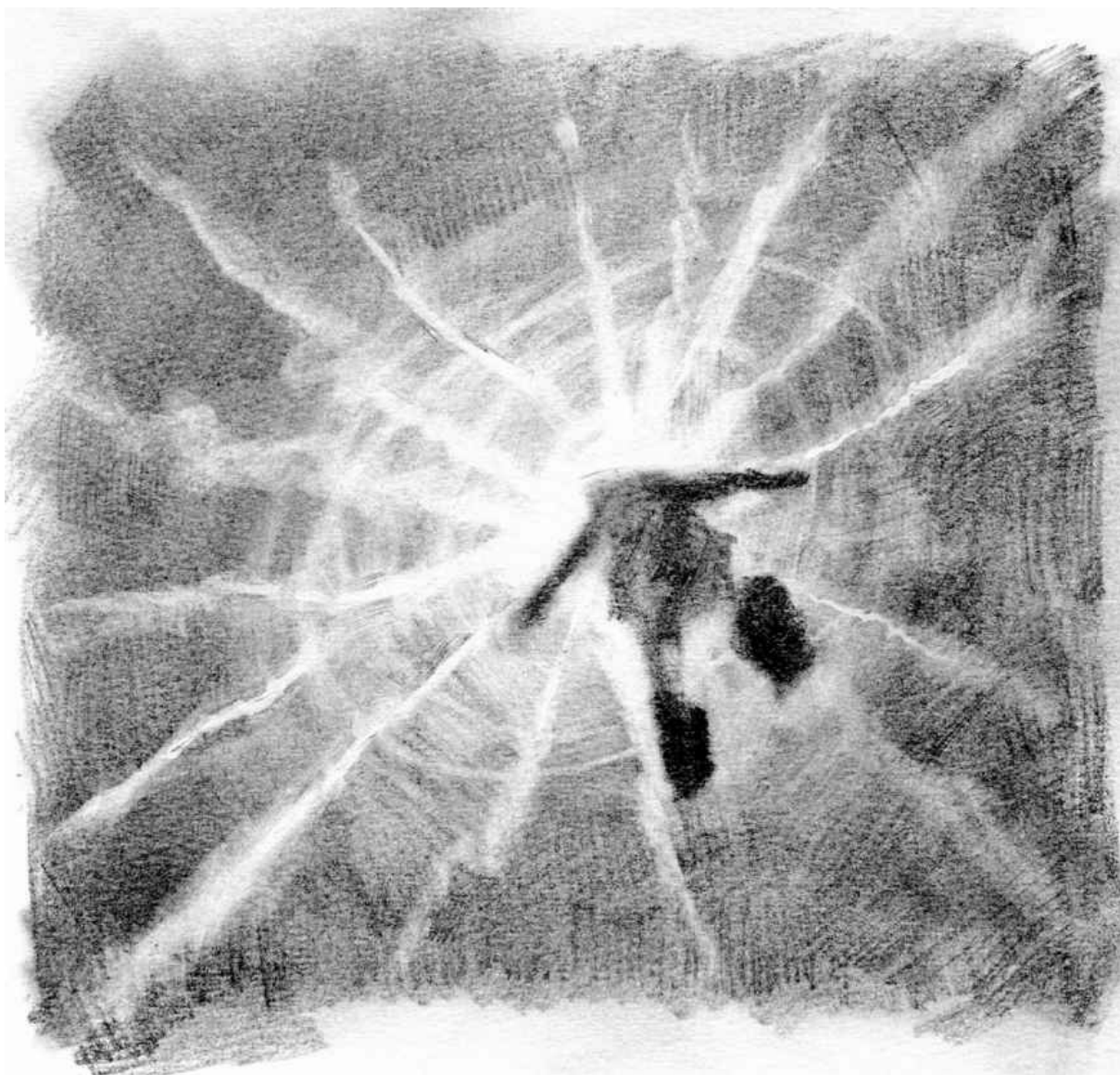
理学家试图通过引进所谓的“自洽性假设”（consistency conjecture）来协调时间的非单向性与因果律之间的矛盾。不过从目前的研究结果来看，这种“自洽性”的一种很有可能的体现方式就是物理规律自动阻止闭合类时曲线的出现。

[3]有人也许会问，如果“创造”虫洞是不可能的，那么所谓“已经存在”的虫洞从何而来呢？这是一个很有趣的问题，我们都知道能量守恒是物理学上的一个基本定律，也就是说物质是不能无中生有的，那么宇宙中的物质从何而来呢？这两个问题有相似之处，由于我们对于宇宙本身的由来还知之甚少，因此这些问题都还没有答案。我们把宇宙中“已经存在”虫洞作为这一节的出发点，不仅仅是把它作为一种可能性来看待，同时也是考虑到“创造”虫洞未必真的已被物理定律所严格排除。在这种情况下，假定存在虫洞（不论其来源），考虑如何将之改造并维持为可穿越虫洞是一个不无意义的问题。

[4]这里所说的质量是“惯性质量”（inertial mass），另外还有一类所谓的“引力质量”（gravitational mass）。在广义相对论中，这两类质量是相等的。另外在相对论中质量是能量的一种，因此本文对负质量和负能量不作区分。

[5]这里的半径是指周长除以 2π 。（请读者想一想为什么要作这个注释？）

[6]有关虫洞的深入分析，以及其他一些值得讨论的方面，比如虫洞与时间旅行之间的关系，量子辐射效应对虫洞的作用等，可参阅拙作《从奇点到虫洞：广义相对论专题选讲》（清华大学出版社，2013年）。



绘画：张京

时间旅行：科学还是幻想？[\[1\]](#)

一、从《时间机器》讲起

众所周知，迄今为止人类在空间与时间上获得的自由度是很不相同的。我们可以沿空间方向作自由运动，却无法随意驾驭时间。时间就像一条漫漫长河，世间万物仿佛是河里的漂浮物，只能随波逐流。

现实的尽头往往就是幻想的起点。如果时间是一条长河，那么在这长河之中是否能有船只呢？漂浮物只能随波逐流，船只却可以劈波斩浪。如果时间长河中能有船只，我们就可以乘坐这种船只进行时间旅行，既可以窥视未来，也可以重返往昔，说不定还能改变历史。在科幻小说中，这种假想的船只被称为“时间机器”。

有关时间机器最早、最著名的小说是英国科幻作家威尔斯（H. G. Wells）的《时间机器》（*The Time Machine*），发表于1895年。不过，威尔斯并不是最早触及时间旅行这一题材的作家，在他之前已经有许多作家涉足过这一题材，其中甚至包括美国讽刺小说家马克·吐温（Mark Twain），他发表于1889年的《康州美国佬在亚瑟王朝》（*A Connecticut Yankee in King Arthur's Court*）据说是最早涉及逆向时间旅行的小说。但在那些比威尔斯更早的文学作品中，普遍没有使用像时间机器这样一种可以让人选择“目的地”（确切地讲是“目的时间”）的旅行器，并且也极少对时间旅行的机制作哪怕只是科幻意义上的说明。而威

尔斯的《时间机器》在这两方面都是突破性的，它很快引起了读者们的巨大兴趣，并于1960及2002年两度被拍成电影，英国甚至为《时间机器》出版100周年发行过纪念邮票。

威尔斯写作《时间机器》的时候，爱因斯坦（Albert Einstein）的相对论尚未被提出，人们对时空的理解大体上还停留在牛顿（Isaac Newton）的绝对时空观上^[2]。但威尔斯却在《时间机器》一书中令人吃惊地提出了将时间作为第四维的观点，与十年后到来的相对论时空观作了戏剧性的遥相呼应。

威尔斯将时间视为第四维，目的是要通过将时间与空间类比为时间旅行开绿灯。那么现代物理学认可这个绿灯吗？这就是本文所要讨论的内容。

二、面向未来与重返过去

我们知道，在牛顿的绝对时空观里，时间和空间不受任何物质及运动的影响（这是“绝对”的主要含义所在）。很明显，在这样的时空观里，时间旅行不具有理论基础，它的存在只是一种幻想。但是狭义相对论的提出对时空观产生了一次重大变革。在狭义相对论中，时间和空间不再是绝对的概念，而是与参照系的选择密切相关。特别是，在运动参照系中时间的流逝会变慢，这是著名的时间延缓效应，它的存在已经被大量物理实验所证实。狭义相对论所带来的这种新结果，为时间旅行开启了第一种具有理论依据的可能性：那就是面向未来的时间旅行成为了可能。

按照狭义相对论，如果有人想要到未来去旅行，他所需要的时间机器就是一艘能以接近光速的高速度运行的飞船。想要到达的未来越遥远，飞船所需达到的速度就越高。如果他想在20年（飞船上的时间）的飞行之后到达两万年（地球上的时间）后的地球上，他所要做的就是让飞船以相当于光速99.99995%的速度飞行10年，然后以相同的速度往回飞。那么20年后，当他回到地球上时，地球上的日历已经翻过了整整两万年，他可以如愿以偿地看到两万年后的人类社会（如果那时候人类社会还存在的话）。可以想象，这样一位来自远古的旅行家将会受到未来的历史学家和考古学家们何等热烈的欢迎。

事实上，不仅未来的历史学家和考古学家将会非常欢迎这样的时间旅行家，与这位时间旅行家同时代的人又何尝不希望他能把自己看到的未来世界的情形带回给大家呢？可惜的是，狭义相对论为面向未来的时间旅行开启了大门，却没能重返过去的时间旅行提供同样的理论可行

性。如果一定要对狭义相对论的数学框架做广义诠释的话，那么只有超光速的运动才可能导致某一类参照系中的时序被颠倒。但是狭义相对论本身在亚光速与超光速之间设置了一个光速壁垒，没有任何已知的物理过程能够使原本亚光速运动的物体——包括人——进入超光速运动状态。因此在狭义相对论的理论框架内，时间旅行家可以到达未来，但却不能重返过去，这与我们在空间中自由自在的运动相比，显然是差得很远的。而且，面向未来的时间旅行不一定需要时间机器才能做到，通过将旅行者冷冻若干年再解冻的手段也可以达到同样的目的。因此时间机器如果存在的话，它真正独特的价值不在于面向未来，而在于重返过去。

那么重返过去的路在哪里呢？

在狭义相对论之后又过了10年，爱因斯坦提出了广义相对论。在广义相对论中，时间和空间不仅如狭义相对论中一样与参照系的选择密切相关，而且还有赖于物质的分布和运动。由此产生的一个不同于狭义相对论的重要结果是：我们对“未来”的定义不再是绝对的了，它会受到物质运动的影响。在不同时刻、不同地点，“未来”有可能指向不同的方向。这是一个奇妙的结果，它表明时空在某种意义上就像流体一样会受到物质运动的拖曳，甚至连时间的方向都有可能因拖曳而改变。

既然时间的方向可以被物质的运动所拖曳，那么有没有可能存在某种物质的分布与运动，它对时间方向的拖曳如此显著，以至于把未来方向拖曳成过去方向，甚至让不同的时间方向首尾相接，连成一条闭合曲线呢？这样的闭合曲线如果存在，无疑就是一种时间机器。因为沿这种曲线运动的飞船每时每刻都在做正常的飞行，感受到正向的时间流逝，但它的轨迹却不仅在空间上，而且会在时间上回到出发点。如果你乘坐飞船沿这样的曲线做一次为期10年的旅行^[3]，那么在旅行结束时你不仅

会回到飞船出发的地方，并且会遇见10年前整装待发的自己^[4]！物理学家们把这种奇妙的曲线称为“闭合类时曲线”，它是时间机器这一科幻术语在广义相对论中的代名词。倘若存在闭合类时曲线，时间旅行就有了理论上的可能性。

那么在广义相对论中，是否存在闭合类时曲线？或者确切地说，是否存在使闭合类时曲线成为可能的物质分布与运动呢？对这个问题，物理学家们做了许多研究。

三、广义相对论与时间旅行

1949年，著名逻辑学家哥德尔（Kurt Gödel）在广义相对论中发现了一个非常奇特的解，描述一个如今被称为“哥德尔宇宙”（Gödel universe）的整体旋转的宇宙。在这种宇宙中，物质的旋转对时间方向会产生拖曳作用，离旋转中心越远，拖曳作用就越显著。在足够远的地方，拖曳作用足以形成闭合类时曲线。因此，在哥德尔宇宙中只要让飞船沿某些远离旋转中心的轨道运动，原则上就可以实现时间旅行。哥德尔这位曾经以哥德尔不完全性定理（Gödel's incompleteness theorems）震撼整个数学界的逻辑学家，又用他的旋转宇宙震动了包括爱因斯坦本人在内的许多物理学家。

可惜的是，哥德尔宇宙并不符合天文观测。首先，我们所生活的宇宙并不存在整体的旋转^[5]；其次，在哥德尔宇宙中宇宙学常数是负的，而我们观测到的宇宙学常数却是正的。因此我们所生活的宇宙显然不是哥德尔宇宙。不仅如此，定量的计算还表明，即便我们真的生活在一个哥德尔宇宙中，也很难实现时间旅行，因为沿哥德尔宇宙中的闭合类时曲线运行一周所需的时间与宇宙的物质密度有关，对于我们所观测到的物质密度而言，沿闭合类时曲线运行一周起码需要几百亿年的时间。因此哥德尔宇宙对于时间旅行并无现实意义。

不过，哥德尔宇宙虽然没有现实意义，但它的发现表明广义相对论的确允许闭合类时曲线的存在，这本身就是一个鼓舞人心的结果。自那以后，物理学家们在广义相对论中又陆续发现了其他一些允许闭合类时曲线的解。比如1974年，美国图兰大学（Tulane University）的物理学家梯普勒（Frank J. Tipler）研究了一个无限长的旋转柱体外部的时

空^[6]，结果发现只要旋转速度足够快，这样的柱体对外部时空所起的拖曳作用也足以形成闭合类时曲线。又比如1991年，普林斯顿大学的天体物理学家高特（John Richard Gott III）发现两条无限长的平行宇宙弦以接近光速的速度彼此擦肩而过时，也会在周围形成闭合类时曲线。与梯普勒人为引进的旋转柱体不同的是，宇宙弦的存在虽然还没有明确的实验证据，但它是许多前沿物理理论所预言的东西。因此高特的结果可以算是把时间机器在理论上的可能性又推进了一步。

但是梯普勒与高特为了数学上的便利都引进了无限长的物质分布（即“无限长的旋转柱体”和“无限长的平行宇宙弦”），这在现实世界中显然是不可能严格实现的。假如物质的分布不是无限的，还可以得到类似的结果吗？物理学家们对此也做了研究，但情况不容乐观：1992年，著名物理学家霍金（Stephen Hawking）给出了一个令人沮丧的结果，那就是如果能量密度处处非负，那么试图在任何有限时空区域内建造时间机器的努力要想成功，都必须产生物理学家们最不想看到的东西——时空奇点^[7]。时空奇点对于研究广义相对论的人来说是并不陌生的，它具有一系列令人头疼的性质，比如物质的密度发散，时空的曲率发散，等等^[8]。虽然没有人确切知道时空奇点的出现会对时间旅行产生什么影响，但这种影响很可能是凶多吉少的。

霍金的这个结果对于建造时间机器无疑是坏消息，但细心的读者也许注意到了，这个结果中有一个限制条件，那就是“能量密度处处非负”。这个条件粗看起来是非常合理的，但我们在介绍虫洞的时候已经提到过，负能量物质的存在不仅在理论上是可能的，而且已经得到了实验的证实。

既然负能量物质可以存在，那么霍金的结果（确切地说是其中的结论部分）就有可能被避免。这方面的研究事实上早在霍金的结果出现之

前就已经有人进行了——当然目的不是为了避免当时尚未出现的霍金的结果：加州理工大学的物理学家索恩（Kip Thorne）与学生莫里斯（Mike Morris）等人在1988年发表的一项有关“可穿越虫洞”（traversable wormhole）的研究中，发现虫洞不仅是空间旅行的通道，而且还可以作为时间旅行的工具——只要让虫洞的出入口以接近光速的速度作适当的运动，就可以将虫洞转变成时间机器^[9]。由于虫洞中含有负能量物质，因此他们这种时间机器可以避免霍金的结果，不导致时空奇点（从这个意义上讲，负能量物质还真是很有“正能量”）。索恩等人的这一研究把科幻小说中最具魅力的两个概念——虫洞与时间机器——联系在了一起，集“万千宠爱”于一身，很快就成为了建造时间机器的热门方案。

但是，索恩等人的虫洞时间机器虽然可以避免霍金的结果，却立即遇到了另一个棘手的问题，那就是虫洞一旦成为时间机器，在类时曲线闭合的一刹那，任何微小的量子涨落都有可能通过那样的虫洞返回过去，与它本身相叠加。这种叠加过程可以在零时间内重复无穷多次，由此产生的自激效应足以在瞬间将时间机器彻底摧毁！这种效应不仅危及索恩等人的“虫洞时间机器”，对其他类型的时间机器也同样具有威胁。1992年，霍金干脆提出了著名的时序保护假设（chronology protection conjecture），认为自然定律不会允许建造时间机器。不过迄今为止，这还只是一个假设，而且霍金的论据也不是无懈可击的，对时间机器的理论可行性持乐观看法的物理学家们陆续提出了一些模型来突破霍金对时间机器的封杀。这方面的讨论目前仍在继续。

四、时间旅行与因果佯谬

有关时间机器的讨论除了探讨它的理论可行性外，还有一个非常重要的方面，那就是探讨时间机器假如存在，我们能用它来做什么？

粗看起来，这似乎不成之为问题，既然能够做时间旅行，那么到达目的时间之后自然应该是想做什么就可以做什么——只要不违反物理学定律。但细想一下，事情又不那么简单。举个例子来说，倘若时间旅行者回到自己出生之前，他能够阻止自己父母的相识吗？这似乎不需要违反任何物理学定律。比如时间旅行者若在自己的父母相识之前，向后来会成为自己父亲的那个人开枪，子弹似乎完全可以在不违反任何物理学定律的情况下击中目标，造成致命伤害。但如果那样的行动成功了，我们就会立刻陷入所谓的“因果佯谬”（causality paradox）之中。因为如果时间旅行者的父母因为他的阻挠而没有相识，那么世上就不会有他；而世上如果没有他，他又如何能够返回过去并阻止自己父母的相识呢？像这样的佯谬在考虑时间旅行时数不胜数，它们都起源于时间旅行对因果时序可能造成的破坏。

这类佯谬该如何解决呢？在科幻小说或电影中，解决的方式往往是通过各种巧合。比如前面提到过的威尔斯的《时间机器》在2002年被拍摄成影片时，或许是为了对主人公建造时间机器的动机做出某种说明，导演增添了主人公情人被害，他试图重返过去加以挽救的情节。在那段情节中，主人公想尽办法，却总是顾此失彼，他的情人总会以这样或那样的方式死去。显然，同样的手法也可以用来避免时间旅行者阻止自己的父母相识。比方说当时间旅行者正要采取某种手段阻止父母相识时，不小心踩到一块香蕉皮摔伤住进医院，从而错过了时机^[10]。这样的解

决佯谬的方式被一些物理学家戏称为“香蕉皮机制”（banana peel mechanism）。在“香蕉皮机制”下，时间旅行者看似能够自由行事，但每当其行为将要导致因果佯谬时，总会受到某些看似偶然的因素干扰，致使行为失败。

这种“香蕉皮机制”很适合编写戏剧性的故事情节。但从物理学的角度讲，很难想象物理学定律需要通过如此离奇巧合的方式来解决佯谬[11]。更何况，香蕉皮机制还有一个致命弱点，那就是它往往只着眼于保证一两个核心事件——比如影片《时间机器》中主人公情人的死亡，或者我们所举的例子中时间旅行者父母的相识——的发生不会被时间旅行所改变，却无法兼顾其他事件。比如影片《时间机器》中主人公的情人以不同方式死亡会在当地报纸上留下不同的报道；我们所举的例子中时间旅行者的摔伤住院也会在当地医院中留下相应的记录。这些事件对特定的故事来说并不突出，但从维护因果时序或历史的角度讲却与核心事件有着同等的重要性。事实上，自然界的各种事件之间存在着千丝万缕的联系，任何看似微小的变化，都有可能通过这种联系逐渐演变成重大事件，这一点对混沌理论中的蝴蝶效应（butterfly effect）有所了解的读者想必不会陌生[12]。

除香蕉皮机制外，在一些科幻故事中还可以看到另外一种观点，那就是在一定程度上放弃因果律，以扩大时间旅行者的行动自由。在这种观点下，历史可以近乎随意地被改变，并且改变的结果可以影响到现实世界中的许多事情。科幻影片《频率》（*Frequency*）体现的就是这种观点。在那部影片中，主人公虽然没有直接进行时间旅行，但他通过与30年前去世的父亲建立联络，具备了间接改变历史的能力。在影片中，历史事件的每一次改变都会直接改变30年后的现实世界。比如由于历史事件的改变导致主人公母亲意外死亡，30年后主人公母亲的相片就会从

相框中突然消失。显然，这种观点几乎等于放弃已知的物理学定律，比试图保护现实的香蕉皮机制更为离奇。

五、凝固长河与平行宇宙

像“香蕉皮机制”或放弃因果律这样的做法，虽然也有物理学家表述过，但总体来说，它们与现实物理学定律之间的差距太大，很少有物理学家会在没有足够证据的情况下，对物理学定律做如此剧烈的变动。对物理学家们来说，更感兴趣的问题是：在现有物理学定律的基础上，能否理解或避免由时间旅行所可能导致的因果佯谬？

对于这一问题，物理学家们尚未形成一致的看法。我们在这里向读者介绍两种主要的观点。

第一种观点认为时间和空间是对物理事件的完整标识。因此一旦时间和空间同时确定，物理事件也就完全确定了。从这个意义上讲，如果我们把时间比作一条长河，那它其实是一条凝固的长河，它的每个截面——对应于一个确定时刻所有物理事件的全体——都是固定的，就像电影胶片一样。按照这种观点，历史只能有一个版本，如果时间旅行者能够回到过去，唯一的可能是他原本就存在于过去。这话听起来有点玄妙，用平直一点的话说就是时间旅行者回到过去后所做的一切只能精确地演绎历史上已经存在过的一个人。如果他试图阻止自己父母相识，却不小心踩到香蕉皮摔伤住了院，那么在历史上就的确存在过这样一个人，乘坐奇怪的机器从天而降，很不幸地踩到香蕉皮摔伤住了院，伤愈后又乘坐奇怪的机器离去。换句话说，时间旅行者并不能对历史做分毫的改变，他甚至连历史的旁观者都不是，因为他本身就是历史的一部分。这种观点对于热衷时间旅行的人来说无疑是令人失望的，因为如果一切都是不可改变的，那么时间旅行也就失去了最重要的价值。

幸运的是，第二种看待时间旅行的观点要开放得多，这种观点来源

于美国物理学家艾弗里特（Hugh Everett III）1957年提出的一种奇特的量子力学诠释——多世界诠释（many world interpretation）^[13]。我们知道，量子力学的一个重要特点就是对量子体系进行测量的结果往往是不唯一的。那么，一个具体的测量结果究竟是如何产生的呢？物理学家们提出了许多不同的观点。有些物理学家认为当我们对量子体系做测量时，体系的状态会发生坍缩，我们观测到的测量结果是一个坍缩后的状态。在这种观点中，状态的坍缩是一个不可预测的过程。与之相反，艾弗里特等人的多世界诠释则认为，并不存在这种不可预测的状态坍缩，量子测量的结果是世界分裂为一组平行宇宙。所有量子力学中可能出现的测量结果都是真实存在的，只不过它们分别存在于各自的平行宇宙而非单一世界中。观测者所得到的测量结果，只不过是（她）所在的平行宇宙中的特定结果而已^[14]。如果我们把这种观点运用到时间旅行中，认为时间旅行者不仅跨越时间，而且还跨越不同的平行宇宙，那么所有的佯谬就都迎刃而解了^[15]。比如时间旅行者阻止自己父母的相识就不再成为佯谬，因为所有这一切都发生在一个不同的平行宇宙中。在那个宇宙中他的父母原本就不相识，他自己也原本就不曾出生过。这与阻止父母相识的时间旅行者本人出现在那个宇宙中并不矛盾，因为时间旅行者是来自于另一个平行宇宙的，在那个平行宇宙中他父母依然相识。在这种观点下，每个平行宇宙的历史仍然是唯一的，但是所有物理定律许可的历史都会在某个平行宇宙中得以实现，时间旅行者虽然无法改变任何一个平行宇宙的历史，却可以自由地选择进入哪一个平行宇宙，他不能改变历史，却可以选择历史^[16]。

六、幻想与历史

经过了这些讨论，现在让我们回到本文的标题上来，时间旅行究竟是科学还是幻想？据说索恩与学生发表有关虫洞及时间旅行的论文时，曾经担心被同事们认为是不务正业。但我们在本文中已经看到，在时间旅行这个主题背后有着一系列值得深入研究的物理学课题。事实上，现在的确有一小部分物理学家——其中包括世界顶尖大学的教授——在对这些课题进行认真的研究。这种研究除了试图探讨科幻小说中这些迷人话题的理论可行性外，一个很重要的动机是要探索现有物理学定律的边界，探索在最离奇的情形下物理学定律可以告诉我们什么。从这个意义上讲，时间旅行无疑是一个有着丰富科学内涵的课题。

但是另一方面，从现实可行性上来讲，起码就我们目前所知的物理学定律而言，时间旅行很可能只是一种幻想。我们在前面讨论过许多有可能形成闭合类时曲线的理论模型，撇开它们面临的种种理论难题不论，在那些讨论中我们还忽略了一个很重要的方面，那就是虽然从结构上讲，闭合类时曲线与能让人类使用的时间机器完全类似，但在规模上却有着巨大差异。以索恩等人的虫洞时间机器来说，为了让人类能够使用这种时间机器，虫洞必须是可穿越虫洞。而我们在有关虫洞的介绍中已经看到，建造可穿越虫洞是一件几乎不可能做到的事情，更遑论让虫洞的出入口以接近光速的速度作特定的运动了。因此，索恩的虫洞时间机器无论在理论上是否可能，在现实世界中实现的可能性都是微乎其微的。

限于篇幅，我们有关时间旅行的介绍到这里就告一段落了。十多年前，霍金曾经问过这样一个问题：假如时间旅行是可能的，为什么在我

们周围至今尚未充斥着来自未来世界的时间旅行者呢？这个问题的潜台词是：时间旅行者没有来到我们周围，最有可能的原因是时间旅行在整个时间长河中——也就是永远——都没有实现过。当然，霍金并没有把这样的问题当作是对时间机器的一个认真的理论诘难。不过，他的这个问题还是引起了一些物理学家的思考，并且他们找到了一种可能的回答：即我们目前所知的有可能实现时间旅行的理论模型，有一个很可能具有普适性的共同特点，那就是不允许时间旅行者回到时间机器存在之前的年代。因此，假如公元2500年有人建造出了时间机器，那么时间旅行者只能访问公元2500年之后的年代^[17]，他们永远无法来到我们周围，更无法像一些科幻小说描绘的那样，回到史前时代去捕捉恐龙——那些历史已经或将要无可挽回地被时间长河所吞没，就像美国物理学家格林（Brian Greene）所说的：在时间机器建造成功之前的每一个年代，都将成为我们以及我们的子孙后代永远无法触及的历史。

从这个意义上讲，如果时间旅行是可能的话，早一天建造出时间机器就是多拯救一天历史。

2006年5月18日写于纽约

2014年12月7日最新修订

^[1]本文曾发表于《科幻世界》2006年第7期（科幻世界杂志社出版）。

^[2]在1892至1895年间，荷兰物理学家洛伦兹（Hendrik Lorentz）等人曾在研究电磁理论时提出过一些有别于绝对时空观的假设，但这些假设并未成为主流，后来则被相对论所取代。

^[3]这里“为期10年”指的是飞船上的时间。

[4]事实上，不仅旅行结束时的你会看到10年前的自己，10年前的你在出发时也会看到10年后凯旋归来的自己。假如你在出发时什么都没看到，说明旅程中必定会发生意外，使你无法回到旅行的起点。在这种情况下，你或许应该取消旅行！

[5]当然，这是指在现有的观测精度内没有发现宇宙的整体旋转。另外，有读者可能会问：什么是宇宙的整体旋转？这种旋转是相对于什么来定义的？这类问题可以视为是跟奥地利哲学家马赫（Ernst Mach）的观点，即旋转必须是相对的，一脉相承。不过，尽管爱因斯坦本人曾经推崇过马赫，但广义相对论事实上并不严格遵循马赫的哲学观点。

[6]梯普勒并不是最早研究这一时空的物理学家，早在1937年，荷兰物理学家范斯托克姆（Willem Jacob van Stockum）就曾研究过这一时空，只不过没有像梯普勒那样对其因果特性进行分析。

[7]确切地讲，许多物理学家都得到过类似的结果，霍金的只是其中之一。

[8]奇点的严格定义本身就是广义相对论中一个非常棘手的课题，这里叙述的只是某一类奇点的特性，更详细的叙述可参阅拙作《从奇点到虫洞：广义相对论专题选讲》（清华大学出版社，2013年）。

[9]具体地说，让虫洞成为时间机器所需的最简单的运动是那种使虫洞两个出口之间的外部空间距离迅速改变，而虫洞本身的长度却不改变的运动。产生这种运动并不容易，但在原则上是可以做到的。关于“虫洞时间机器”的更详细介绍，可参阅拙作《从奇点到虫洞：广义相对论专题选讲》（清华大学出版社，2013年）。

[10]当然，这只是最简单的巧合（不过“香蕉皮机制”因之而命名，故特意举出）。为了情节的需要，我们还可以设想更为复杂的巧合。比方说时间旅行者试图向后来会成为他父亲的那个人开枪，却因为心情矛盾导致枪法失准，没有击中“父亲”，却击中了“父亲”的情敌！他试图阻止父母相识的行动非但没有达到目的，反倒为他父母的结合铺平了道路。他的行动不仅没有破坏因果关系，反而成为了维护因果关系所必需的，等等。像这种近乎宿命的巧合在科幻故事中用得也很多。

[11]尽管如此，还是有物理学家做过这方面的考虑。比如俄国物理学家诺维科夫（Igor Novikov）曾经提出过一个假设，认为物理学定律会——哪怕通过离奇巧合的方式——自动保证不出现因果佯谬。这个假设被称为“诺维科夫自洽性假设”（Novikov consistency conjecture），它可以算是香蕉皮机制的理论版本。不过这个假设一直缺乏具体的实现方式。

[12]举个例子来说，如果时间旅行者回到过去后把一块小石头放在路上，然后离开。这样的事件无疑是非常微不足道的，但它有可能导致某位行人因踩到石头而扭伤脚。而这位倒霉的行人有可能恰好是一位物理学家，他正要去做一个有关时间旅行的学术报告，却因为扭伤了脚而取消报告。而那个学术报告的听众中有可能恰好有一位年轻人因为这个报告的影响而投身于时间旅行的研究，并最终成为时间机器的建造者。在这种情况下，时间旅行者放在路上的小石头对历史的影响就扩大成了尖锐的佯谬。因为正是这块石头的出现，使得一位物理学家取消了学术报告，既而又使得一位年轻人因没有听到这个学术报告而不再以时间旅行作为自己的研究方向，而这最终导致了人类没能研制出时间机器。但如果人类没能研制出时间机器，时间旅行者又如何能够放置那块小石头呢？

[13]艾弗里特是多世界诠释的提出者，不过“多世界诠释”这一术语却是美国物理学家德惠特（Bryce DeWitt）提出的。

[14]需要指出的是，多世界诠释的原始表述其实并不依赖于像“多世界”或“平行宇宙”那样的概念。后来流行的“多世界”或“平行宇宙”概念从某种意义上讲是对多世界诠释本身的诠释。

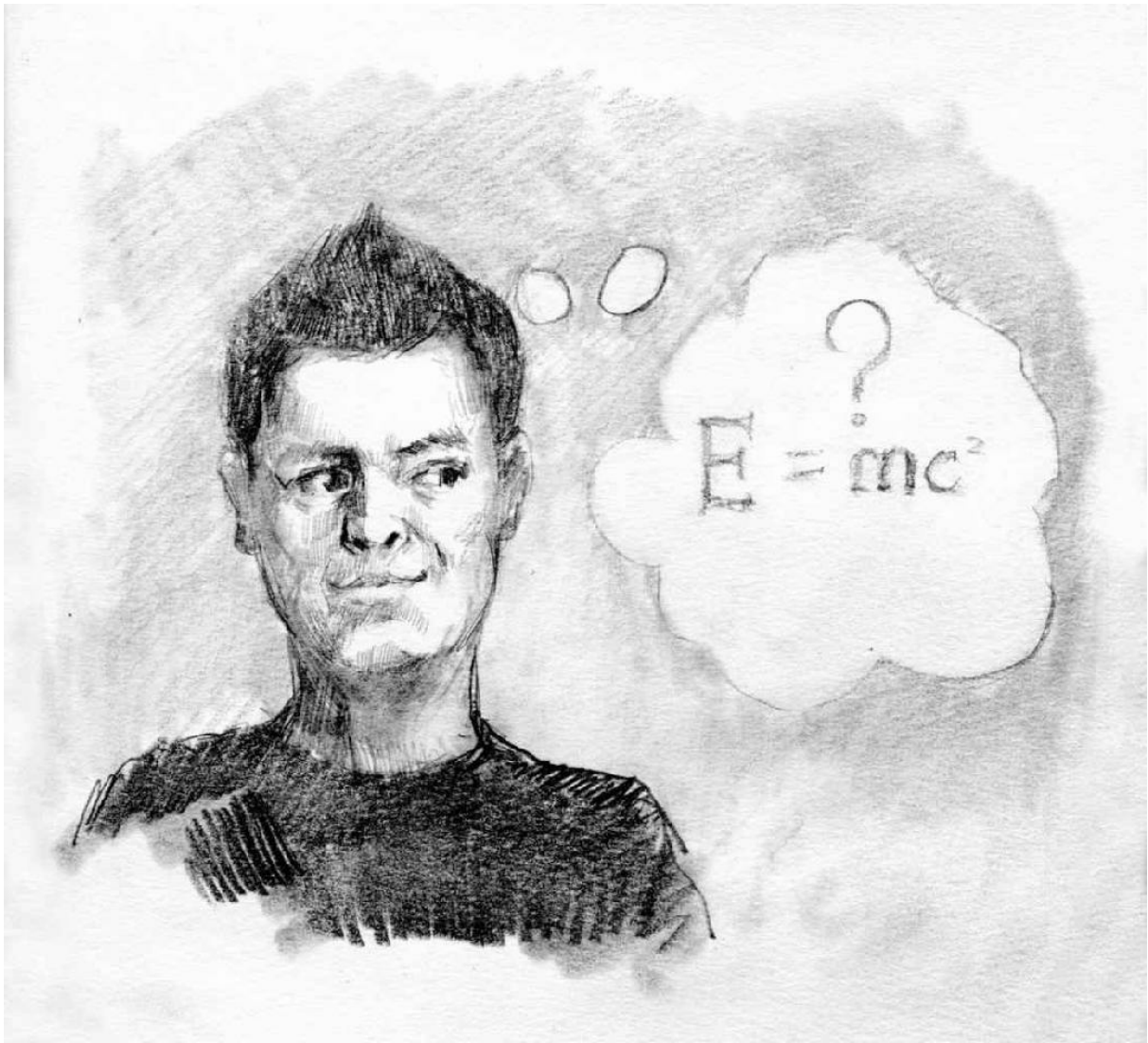
[15]当然，这里所谓的“迎刃而解”，是建立在有着极大争议性的平行宇宙概念之上的，因而本身也是有着极大争议性的。此外，所谓“迎刃而解”，首先还假定所讨论的问题有意义，这同样有可能是不成立的，因为时间旅行完全有可能是如霍金猜测的那样被物理学定律所禁止的，由时间旅行所导致的因果佯谬也因此完全有可能是伪问题。

[16]即便按照这种观点，科幻小说中的许多情节也是不可能实现的。比如通过时间旅行者对某个历史事件的干预来改变人类命运就是不可能的。时间旅行者的努力，只能使他自己进入一个人类命运截然不同的平行宇宙中去，而试图通过这一努力来改变自己命运的原平行宇宙中的其

他人的命运，将不会因此而改变。

[17]注意，这并不是说时间旅行者只能作面向未来的时间旅行。在时间机器存在之后的那些年代之间，他们的旅行既可以面向未来也可以面向过去，他们只是无法回到时间机器建造之前的年代去。

第四部分 其他



绘画：张京

从民间“科学家”看科普的局限性

半年多前，我在网上偶然发现了一个名为“超弦学友论坛”的网站。

那是一个以讨论超弦理论及相关话题为主的中文学术论坛，设有一个主论坛和一个灌水区，后者是留给与学术无关的话题的。常言道：“林子大了，什么鸟都有”，建一个灌水区可以让不做学术的鸟儿也有个试嗓子的地方。与其他论坛相比，“超弦学友论坛”的最大特点，是有几位中科院及中国科技大学的教授主持，因此秩序相对好些。我初次光顾该论坛的时候，主论坛上有教授和同学们的许多讨论，就像一个网络课堂。但不久前旧地重游，却发现“林子”里的光景已经大变，主论坛上有大批民间“科学家”往来穿梭，在灌水区却发现了一位原先很活跃的教授的踪迹。教授在那里发了一个短短的跟帖，所跟的是他本人被别人转过来的一篇文章。教授在跟帖中写道：

谢谢转帖，但我希望尽量不要将我的东西转贴到隔壁，因为隔壁演变成了一个民间“科学”论坛。

这一跟帖对论坛无疑是一个警讯，不久之后论坛的管理员出来删除了一些帖子。

“超弦学友论坛”所遭遇的这种情况在网上是有一定代表性的。互联网的发展给原本需要自费印刷资料、自费前往学校或科研院所推销“理

论”的民间“科学家”们提供了极大的便利，使他们亮相的成本大幅降低，“出镜率”也因此大幅提高。大众对民间“科学家”的态度遂成为近年来较有争议的一个话题。

民间“科学家”这一概念并没有一个很严格的定义，因为这是一个具有相当复杂性的群体。往上了看，一部分科学家在其童年或少年时期的思维形式与某些民间“科学家”也有一定的相似性；往下了看，许多伪科学或反科学人士的思维形式与民间“科学家”同样有一定的相似性。粗略地讲，民间“科学家”主要有这样两条特征：

一、民间“科学家”没有接受过系统的科学训练

这一条几乎是定义性的。多数民间“科学家”自己也坦承这一条，就像在过去某个年代里，大家并不避讳自己的赤贫家境一样。这里所说的系统的科学训练并不单单指的是科班出身，完全也可以是达到同等层次的高水平的自学。此外，这里所说的系统的科学训练是以真正学到手为判据的，而不是仅仅混到一个文凭。

二、民间“科学家”无意接受系统的科学训练

这一条往往被人忽略，不过我觉得这一条其实很关键。因为即使是最优秀的科学家，也并非生来就接受过系统的科学训练的，因此“没有接受过系统的科学训练”并不是区分民间“科学家”与科学家的最本质特征。许多民间“科学家”也常常用科学家在童年或少年时期的故事来为自己辩护。但被民间“科学家”们有意无意地予以忽略的是，他们的思维形式与真正的科学家在童年或少年时期的思维形式虽有一定的相似性，但这种相似性却永远地凝固在了那样一个年龄段上，仿佛自幼年起就停止了发育。民间“科学家”们虽然对科学充满了雄心壮志，试图“研究”科学界最艰深、最宏大的课题，试图“推翻”科学界最有实验基础的理论，但他们数十年如一日的行为却只是在一个极低的水平上循环往复。他们可以花几十年的时间来做“研究”，却无意拿出几年的时间来系统地学习科学。科学界的文献是开放的，但由于他们无意接受系统的科学训练，从而在实质上放弃了阅读和理解科学文献的能力。因此他们的“理论”无论用什么时髦的科学术语来包装，用科学界的标准来衡量，都只是停留在一种十分原始的、伽利略之前的思维水准上。

这两条特征当然既不是完备的，也不是毫无例外的，想要在这样一个模糊的领域中建立一个绝对清晰的定义是一种徒劳。但这两条概括了绝大多数民间“科学家”的基本特征。

远离了系统的科学训练，远离了科学文献，民间“科学家”获取知识的主要来源是科普读物。因此大量民间“科学家”的出现也使我们看到了科普在向大众传播科学知识的过程中所显露出的一个薄弱环节：那就是

科普对于现代科学的通俗化处理具有一定程度的误导性。

这么说让我自己觉得很不安，因为我非常敬重科普，希望这样的说法不会被理解为轻视或贬低科普。我想要通过本文表达的观点是，科普是好东西，但她所面向的读者群体决定了她有无可避免的局限性，她不能作为科学研究的完整背景。一个试图研究科学的人所需获取的基础知识绝不能止步于科普的层次。科普的作用是让没有机会研究科学的人了解科学；让有机会研究科学的人喜欢科学，给他们一个“第一推动力”，让他们超越科普、接受系统的科学训练、继而投身于真正的科学研究。科普不应该起的作用是让有志于研究科学的人以为那就是科学，以为读过科普就算懂得了科学。遗憾的是，科普对民间“科学家”所起的恰恰是它不应该起的作用。

科普在本质上是面向非专业读者的，因此对许多科学概念和理论——尤其是高度抽象的现代科学概念和理论——不得不做极大的简化。这其中最重要的一个简化就是抽去了科学的数学框架，取而代之的是一些文字化的描述以及与日常经验的类比。与这种对科学概念和理论的简化相平行的，是对科学研究过程的简化。科学发现往往被简化成几个概念在科学家脑海里“灵机一动”式的组合。仿佛牛顿的万有引力定律真的就是被苹果砸了脑袋后“灵机一动”就想到了；仿佛爱因斯坦的广义相对论真的就是从几个像“升降梯实验”那样的理想实验中“灵机一动”就得到了。现代科学的研究既有灵感的显现，又有大量扎实而复杂的数学演算及实验，两者相辅而成。但在科普读物中前者给人留下的印象往往远远深于后者，因为前者大体上是概念之旅，既新奇浪漫又富有戏剧性，而后者相形之下不仅显得枯燥乏味，而且往往不是文字叙述所能够完全涵盖的。科普读物的这些局限性都极其明显地体现在民间“科学家”们的“理论”以及他们的“研究”方法上。

什么时候的科学是基本上没有数学结构的呢？那是古代的科学，比如我国古代的五行学说，古希腊的元素学说，等等。在那些学说诞生的年代里，概念和术语的简单组合、纯粹的思辨就可以成为科学（自然哲学）。但是自伽利略之后，科学逐渐脱离纯粹的思辨而进入了以实验和数学体系为主导的时代，现代科学因此而获得了令人赞叹的严密性和精确性。现代科学的这些特点在许多科普读物中都得到了强调，有时甚至是反复的强调。许多科普读物的作者本身就是第一流的科学家，他们深知科学的真谛，他们的科普作品中绝没有忽略现代科学的任何一个重要特征（因此我们讨论的是科普的“局限性”而非“缺陷”）。但现实的情况却是，同样的一部作品对读者所起的作用是和读者本身的知识背景密切相关的。接受过系统科学训练的读者（包括有学术基础的科普作者本人）会自然而然地将科普中的文字叙述与自己在科学训练或研究中的知识及经验相结合，从而获得完整而深入的理解；但对于没有接受过系统科学训练的读者来说，文字化的叙述往往就只会产生文字化的理解。这种理解对于普通读者来说是足够了，但对于一个有志于从事科学研究的人来说却是远远不够的。读100遍“爱因斯坦花了整整 N 年才完成广义相对论”的故事，也远远不如自己动手花 N 个小时来再现一遍爱因斯坦对水星近日点进动值的计算更能体会科学研究的感觉，更能体会现代科学描述自然的方式。这就好比是一个学编程的人，看几本编程的书，却一行程序都不写是学不到编程的精髓的。这种“动手体会”的要求当然不是针对普通读者的，如果是的话也就不需要科普了。但是对于真正有志于从事科学研究的人来说却是必须的。

热衷于砍杀相对论的民间“科学家”们，在挥舞屠刀之前，可否先与现代科学的数学体系做哪怕只是一次这样的“亲密接触”？可否先对人类智慧几百年来的成就做哪怕只是一个细节上的深度了解？

一部分民间“科学家”之所以用自己浅陋不堪的“理论”去挑战现代科学，还往往能挑战得神气十足、老气横秋，乃至盛气凌人，其中很重要的一点就是他们是彻底地“轻装上阵”，他们不仅扔掉了现代科学的数学框架，也扔掉了现代科学背后庞大的实验基础。所以他们可以声称自己的一个没有任何定量结果，没有任何精密实验支持的“理论”超越或推翻了一个有坚实实验基础的科学理论。连科学是人类描述自然的一种努力——从而必须尊重实验观测——这样基本的原则都可以视而不见，现代科学在他们手中自然就变得可以任意宰割了。但是离开了这两者（数学框架和实验基础），科学就退回到了伽利略之前的时代，这事实上也就是绝大多数民间“科学家”所能达到的最高水准（甚至连这样的水准也已经是一种高估，因为哪怕在伽利略之前也已经有不少的学者，比如哥白尼、托勒密等，用相当观测化和数学化的方式来构筑理论了）。民间“科学家”们如果意识不到科普以及他们建立在科普之上的知识体系的局限性，只怕永远也超越不了这一水准。

提出了“统一场论”的民间“科学家”们，可否告诉我们，原则上——也就是不劳您亲自动手，哪怕给个思路也行——如何用你们的理论来推算一个像水星近日点进动值那样的实验结果？

科普并无过错，不仅无过，且有大功。但科普有其局限性。这种局限性只有当她被有志于从事科学研究的人视为科学本身，并以之作为自己“研究”科学的基础时才会显现出来。记得小时候读过一则古老的哲学故事，说有一群人居住在山洞里，面向石壁、背朝洞口。在日月星光的更替中，他们可以看到外部世界在石壁上的投影，于是他们研究起了投影的运动，日复一日，年复一年。但他们谁也没有转过身去看一眼山洞外的世界，他们一直以为那些投影就是整个的世界。科普就好比是那些石壁上的投影，她是科学的一组影像，而民间“科学家”们则好比是山洞

里的那群人，他们在研究影像。

影像没有错，但它有局限性，研究影像也没有错，但如果认为影像就是整个的世界，那就错了。

2003年7月2日写于纽约

什么是民间“科学家”

One of the symptoms of an approaching nervous breakdown is the belief that one's work is terribly important.

Bertrand Russell, 1930

一、新民科引发的问题

2003年7月，我曾写过一篇有关民间“科学家”（简称民科）的文章：《从民间“科学家”看科普的局限性》^[1]。在那篇旧作中，我归纳了民间“科学家”的两条主要特征：

（1）民间“科学家”没有接受过系统的科学训练。

（2）民间“科学家”无意接受系统的科学训练。

那篇旧作由于发表较早，在本站的同类文章中影响较大，被包括维基百科“民科”词条在内的很多网站引用。不过自那篇旧作发表以来，我逐渐意识到它所归纳的民科特性过于狭窄，只适用于早年常见的传统民科。这些年来，我接触到了很多新类型的民科，他们与传统民科有一个很大的区别，那就是带有“教授”、“研究员”、“博导”等学术头衔。当然，在腐败大潮席卷神州的今天，那些头衔不一定都货真价实（确切地说“价”可能是实的，但“货”不一定真）。但不可否认的是，也确实有一些民科是或者曾经是——以后者居多——货真价实的“教授”、“研究员”、“博导”等。那些人都曾受过系统的科学训练，从而并不符合那篇旧作所归纳的民科特征。但那些人的所作所为却与传统民科并无二致，即通过非学术渠道发布不被学术界接受的“论文”，宣称自己破解了重大科学难题，或推翻了重大科学理论^[2]。

那些新民科的涌现，使我有必要重新讨论这样一个问题：什么是民间“科学家”？

二、有关民科的几个较具误导性或典型性的观点

我之所以要讨论这个问题，除了想弥补旧作的不足外，还有一个用意是想借讨论这个问题之机，顺便澄清一些有关民科的较具误导性或典型性的观点。那些观点大都来自过去这些年我接触到的民科及其同情者。若无意外，我希望本文成为我最后一篇有关民科的独自成篇之作（因为多写此类文章并无太大价值，反而会让我“日进斗敌”）。对过去接触到的与民科有关的较具误导性或典型性的观点一并做些分析，可以避免留下太多有可能使我旧话重提的由头。具体地说，本文将分析以下三种较具误导性或典型性的观点：

（1）“泛民科”观点。持这种观点的人认为民科这个概念是相对的，将别人视为民科的人（比如在下），在更高水平的人（比如诺贝尔奖得主）面前，自己也将被归为民科。这种“泛民科”观点的威力是巨大的，它让我想起很多年前看过的一部美国喜剧系列片《火星叔叔马丁》（*My Favorite Martian*）。在那部系列片中，火星人“马丁叔叔”的头顶可以升出一对具有隐形功效的天线，但有一次那天线出了故障，升起之后没能使“马丁叔叔”隐形。这下麻烦大了，因为那会暴露“马丁叔叔”的火星人身份。怎么办呢？“马丁叔叔”想出了一个高招，那就是让那种天线成为流行饰品。一旦大家都戴上那样的天线，“马丁叔叔”的天线就不再扎眼了。这种让所有人都变得相似的方法成为了保护“马丁叔叔”的最佳方法，用“马丁叔叔”自己的话说（大意）：把一棵树藏起来的最好办法就是把它藏在树林里。“泛民科”观点对民科所起的作用也是如此，它通过让所有人都变成程度不同的民科，而让真正的民科得以遁形（当然，这或许只是民科同情者们的一厢情愿，民科自己恐怕非但不想遁形，反而急切地想要展示自己独有的“天线”）。

(2) 将民科与学术界的非主流研究相提并论的观点。众所周知，学术界的研究有许多类型，其中既有主流，也有非主流。非主流研究的存在对学术界是有价值的，不仅因为它们中的某些或许有朝一日会变成主流，或具有部分价值，而且也因为它们与主流研究的竞争有时能帮助揭示主流研究的不足之处，或促使主流研究者将自己的理论表述得更严密。但非主流研究按定义就意味着职位及同路人较少，从而在谋职、发表等方面带有一定的弱势性，这一点往往被民科引为同类。将民科与学术界的非主流研究相提并论，可以起到模糊民科与学术界界限的作用，从而间接提高民科群体的地位。

(3) 对“民科”中的“民”字作字面解读的观点。持这种观点的人认为所谓民科，就是栖身“民间”的科学家，而学术界则是所谓的“官科”（因为“官”与“民”相对）。如果说将民科与学术界的非主流研究相提并论可以起到模糊民科与学术界界限的作用，那么将学术界视为与“民科”相对的“官科”所起的作用则恰好相反，那就是使民科与学术界划清界限，并对后者进行抹黑。因为在中国，“官”字所代表的形象是相当负面的，如果学术界跟“官”是一丘之貉，那么很多人也许会出于对“官”的反感而宁愿支持“民”科。除此之外，将学术界视为“官科”还有一个好处，那就是便于民科用阴谋论的手法为自己的受迫害情结寻找依据，即把自己打扮成被“官”欺压的“民”，把自己的观点不被学术界接受说成是自己的创见被后者所打压（他们显然没有意识到，科学史上有无数比他们新颖百倍的创见都被学术界接受了）。

三、民科的定义

以上三种观点，是我这些年接触到的有关民科的观点中较具误导性或典型性的。要想澄清这些观点，有必要对民科这一概念做一个适当的定义。这个定义的思路在本文开头其实已经涉及到了，那就是从民科的行为及发布渠道入手。当我们把某些带有学术头衔的人列为民科时，所依据的正是他们的行为及发布渠道与传统民科相同。由此可见，从这一角度入手定义民科要比我那篇旧作所列举的背景特征更具适用性。不仅如此，从这一角度入手也比列举民科的其他特征，比如狂妄、偏执等，更具适用性，因为后者往往与民科的具体个性有关，不易一概而论，而且那些特征大都具有贬义，容易引起不必要的意气之争。有鉴于此，本文拟从行为及发布渠道入手，引进以下定义：

所谓“民间科学家”（简称民科），是指以非学术渠道为主，宣称推翻重大科学理论，或破解重大科学难题的成年人。

在应用这个定义前，让我们对定义中的若干用语作一些简短说明：

- “非学术渠道”是指除学术刊物、学术机构预印本、学术会议等正规学术成果发布渠道以外的其他渠道。其中目前最受民科青睐的是博客、论坛、垃圾邮件等渠道^[3]。
- “以非学术渠道为主”中的“为主”二字，是考虑到托学术腐败的福或单凭运气，民科们有时也能在学术刊物上发布“论文”，从而不宜一刀切。不过由于能被民科渗透的刊物通常水平较低，加上民科“论文”的水平更低，发表之后势必石沉大海，难以彰显“鸿鹄之志”，因此民科不管“论文”发表与否，都会以非学术渠道为主进行长期推销，以扩大影响^[4]。这“为主”二字的另一个作用，则是防止有人以

某些科学家也撰写博客或参与论坛活动为由，来混淆其与民科的区别。对后者来说，撰写博客或参与论坛活动并非发布论文、谋求承认的主要渠道。此外还要说明的是，这“为主”二字因涉及不同渠道间的比较，有时需要一定的时间才能做出可靠的判断（一般来说，民科通过非学术渠道对自己“论文”所做的推销越卖力，就越便于人们作出可靠判断）。

- 本定义所说的“科学”既包括自然科学（物理、天文等），也包括数学。
- 本定义所说的“重大科学理论”既包括意义或影响重大的理论（比如相对论、量子力学等），也包括其他具有坚实基础——从而往往能“牵一发动全身”——的命题、定理等（比如“尺规化圆为方的不可能性”等）。
- 本定义所说的“重大科学难题”既包括未解决的难题（如哥德巴赫猜想、黎曼猜想等），也包括已解决的难题（如四色定理、费马大定理等），因为重新“破解”后者也是民科们所热衷的。
- “成年人”三个字的使用，是为了避免将尚在系统求学阶段的年轻人列为民科。如我在旧作中所说，民科的某些特征与童年或少年时期的科学家有一定的相似之处，民科们时常利用这一点为自己辩护。一个合理的民科定义则必须将这种混淆排除在外^[5]。

四、民科定义的应用

定义既已给出，我们就可以用它来分析一些东西了。

首先可以看到的是，上述定义与我旧作中所归纳的传统民科的两条特征是相容的（但涵盖面更广，因为它还涵盖了本文开头所提到的带有学术头衔的民科）。因为满足那两条特征的传统民科显然无法跻身学术界，从而必然只能以非学术渠道为主来宣布自己的“发现”。这表明传统民科符合上述定义。其次我们还可以看到，民科的若干常见言论与上述定义也有很好的相容性，甚至有一定的因果传承关系。比如正因为民科是以非学术渠道为主宣布自己的重大“发现”，从而往往要面对如此重大的“发现”为何要用如此“简朴”的渠道发布的问题，对此的“最佳回答”莫过于是把自己比喻成当代的哥白尼、布鲁诺，把学术界比喻成当年的教廷（或当今神州的官场），这正是民科言论中很常见的类型。而一些民科言论所展现出的病态的狂妄与偏执，则与自以为作出重大“发现”后成名欲的爆棚，及在学术渠道前“小扣（或猛踢）柴扉久不开”后的愤恨不无关系。

接下来让我们再用上述定义来分析一下前面提到的那几种具有误导性或典型性的观点：

（1）“泛民科”观点。这种观点的谬误之处在于忽略了上述定义中的“以非学术渠道为主”及“破解重大科学难题”、“推翻重大科学理论”等界定。一个人是否是民科并不单纯取决于水平高低，即便要论水平，也应该论相对于自己研究目标而言的水平。一个有一定水平的人若从事的是自己水平不能及的“研究”（比如“破解重大科学难题”或“推翻重大科学理论”）而至偏执的程度（即无法以学术渠道为主进行发布却仍

执迷不悟），他就会成为民科；而一位中学物理教师如果从事的是自己的教学研究，他就不是民科^[6]。

（2）将民科与学术界的非主流研究相提并论的观点。这种观点的谬误之处在于忽略了上述定义中的“以非学术渠道为主”这一界定。学术界的非主流研究与主流研究一样，都是以学术渠道为主发布成果的。一旦离开那样的渠道，它们就不再是学术界的非主流研究了。只有在那时，它们才会与民科有可比性（可惜那时它们对提升民科群体的地位往往已不起作用了）。

（3）对“民科”中的“民”字作字面解读的观点。这种将民科理解为栖身“民间”的科学家，将学术界定义为“官科”的观点同样不符合上述定义。因为上述定义丝毫未涉及人在哪里的问题，它所关注的只是行为及发布渠道。一个身在民间的研究者如果以学术渠道为主发布研究成果，接受同行评议，他就不是民科（一个最典型的例子就是常被民科们引为“知己”的尚在专利局时的爱因斯坦）；反过来，一个身在学术界甚至有过杰出成就的人若只能以非学术渠道为主来宣称重大“研究”，那么无论他身在何处，名声是否显赫，起码在该项“研究”中的表现可被视为民科（带有学术头衔的民科就属于此类）。如果一定要对民科中的“民”字作一个字面解读的话，那么虽然绝大多数民科确实身在民间，这个字的本质含义却应该界定为发布渠道的民间性。

在本文最后有必要指出的是，如我在旧作中曾经说过的，对民科这样一个概念做任何定义或归纳都不可能做到完备或精确。本定义也不例外，除有可能存在反例或难以判别的个例外，其涵盖面也还不够广（虽比旧作来得广，却仍不足以涵盖全体）。比如由于将发布渠道作为定义的一部分，使得正在从事“研究”，但尚未发布任何消息（从而与外部社会尚处于绝缘状态）的人无论其“研究”多么民科化，都不在本定义的涵

盖范围之内；又比如由于将“宣称推翻重大科学理论，或破解重大科学难题”作为定义的一部分，使得“胃口”小，不以之为目标的人无论其“研究”多么民科化，也并不在本定义的涵盖范围之内^[7]。

2011年3月5日写于纽约

2014年1月9日最新修订

[1]已收录于本书。

[2]需要说明的是，那些曾经受过系统科学训练的民科在行为模式上与传统民科还是有一定区别的，主要体现在他们不像传统民科那样“无知者无畏”，他们文章的措辞要比传统民科来得谨慎，语气不像后者那样斩钉截铁。

[3]这里用“垃圾邮件”一词，并非刻意贬低，因为“垃圾邮件”是指未经对方许可强行发到用户邮箱中的邮件（unsolicited mails），尤其是指同时发给多个用户的邮件（unsolicited bulk mails）。民科以邮件方式向别人发送“论文”时所发的往往正是符合此定义的邮件。

[4]顺便说一下，这一行为隐含着民科的“论文”无论发表与否，都未被学术界真正接受，以及民科对自己“成就”进行反复宣称等未在定义中直接列出的特点。

[5]另外可以补充的是，这里的“成年人”一词只是简略说法，并不等同于年龄意义上的成年人，由于它的作用是避免将尚在系统求学阶段的年轻人列为民科，因此其含义也是以是否仍处于系统求学阶段为界定的。一个年龄意义上的未成年人若在从事本定义所述的民科行为的同时，已不再接受系统的科学训练，那对于本定义来说就可被列为“成年人”。

[6]打个比方来说：小蛇虽小，若吃的是小动物，那就是正常行为；大蛇虽大，若意在吞象，且不死不休，那就是民科行为。

[7]不过，我见过的民科不少，那样的人却尚未见过，这或许并非偶然，而是因为“胃口”小，甘心做小课题，不好高骛远的人不容易成为民科。

学物理能做什么？[\[1\]](#)

说实话，接到这篇让我向年轻人介绍“学物理能做什么？”的约稿时，我的第一反应是婉拒。当然不是怕“年轻人”三个字把自己衬老了，而是觉得以我已经转行了的身份来写这样的文章，恐怕会适得其反。因为这篇约稿的背景，是物理在高考志愿中逐渐受到冷落，而约稿的目的，则是要鼓励年轻人选择物理。对于这个目的来说，我恐怕是一个坏榜样。不过约稿编辑洞察先机，在约稿信中直接把我归为“工作转行，却并没有真正离开物理”这样一类人的代表，断了我的托词。于是我只好老老实实来写这篇文章。

我体会编辑让包括我在内已经转行的人也来写这个话题，是想让年轻人知道，即便他们今后实际从事的是别的职业，也依然可以报考物理专业。因为他们在这一专业所受的训练，对从事别的职业同样会有助益，甚至会有独特的优势。这样的意思我在以前的文章中曾经作为体会述及过，但从未当作一种专业选择的策略向任何人推荐过，因为在我看来，物理所具有的这种优势是不能当作策略来用的。任何人如果出于喜爱物理以外的其他动机而选择物理，其结果很可能是既学不好物理，也无法实现原本希望通过物理来实现的其他目标。因为物理对于不喜爱她的人来说，并不是一门容易的专业。

但另一方面，物理在高考志愿中所受的冷落，未必是因为越来越多的年轻人已不再喜爱物理，而很可能只是因为年轻人变得更现实了，或

受到了来自亲朋好友更现实的劝告。我想本文的真正读者应该是这部分年轻人，而本文所要表述的观点是：请不要因为担心未来的出路而放弃自己喜爱的物理。这并非是劝诫任何人为了理想放弃现实，而只是说，起码就物理而言，这两者之间的距离并不像许多人以为的那样遥远，从而没有必要担心，更没有必要因此而早早地放弃自己的理想。一个人源自年轻时代的激情，在未来的人生之路上往往是难以再现的，给自己一个机会去追求并真正了解自己的兴趣，是明智而无悔的选择，过早地放弃——尤其是建立在错误理由之上的放弃——则是令人惋惜的。

好了，现在我们言归正传，从求职的角度来说说“学物理能做什么？”。其实这个问题基本上是不需要回答的，因为相反的问题——即学物理不能做什么——恐怕反而是比较困难的。在学物理所能做的事情当中，除了物理本身以外，还涉及许许多多其他职业，本文只举其中一个例子：金融。之所以举这个例子，除了金融是一种热门职业外，更重要的是因为这个曾经与物理风马牛不相及的职业，比其他职业更能体现出人们从学物理中获得的能力所具有的广泛适用性。

如果不考虑零星的个例，物理学家进入金融界大致可以追溯到20世纪70年代末的美国。当时由苏联发射人造卫星在美国引发的科技震荡及热潮已渐渐消退，很多物理专业的学生开始寻找新的求职领域。而在那之前不久，金融领域本身发生的一些变化，恰好为物理学家的进入创造了条件。1973年，当时在芝加哥大学（University of Chicago）和麻省理工学院（MIT）的经济学家布莱克（Fischer Black, 1938—1995年）、斯科尔斯（Myron Scholes, 1941—）及默顿（Robert C. Merton, 1944—）等人提出了有关金融衍生品（Derivatives）的数学模型。这个数学模型（称为布莱克-斯科尔斯模型）的基础是一组偏微分方程，而这组偏微分方程与物理学上用来模拟随机过程的某些方程式具有一定的相似

性。显然，物理学家们在研究这种方程式上具有很大的优势。而且这种优势不仅仅来自于那些方程式与物理方程式之间的相似性，更多地是来自物理学家们所具有的处理包括那种方程式在内的各种复杂问题的普遍技巧，以及修正旧模型、构建新模型的能力。在瞬息万变的金融世界里，这种能力无疑具有极大的重要性。

金融衍生品在20世纪70年代时还是一种不太重要的东西，默顿在当年论文的开头甚至表示，为此发展一套理论也许是不值得的。但在30多年后的今天，金融衍生品的市场规模却远远超过了像股票那样的传统金融产品。1997年，默顿和斯科尔斯因为当年那“也许是不值得的”工作获得了诺贝尔经济学奖（布莱克很遗憾地因为已经去世，无法分享这一荣誉）。而物理学家参与其中共同打造的这种以金融模型分析为主要职责的新角色，也早已成为了金融界的一种重要的新兴职业：定量分析师（quantitative analyst，简称quant）。由于这一职业的兴起，在20世纪90年代，华尔街成为了向物理学家提供职位最多的领域之一。在某些公司中，物理学博士的人数竟然占到了公司总人数的三分之一甚至更多。到了2007年，就连物理学界最著名的论文预印本档案馆arXiv.org也为参与金融分析的物理学家们增添了一个新的论文类别：定量金融（quantitative finance）。这个类别如今每个月都有几十篇论文。

在物理学家眼里，一个领域的成熟往往意味着它的淡出。对于金融分析来说，这一天即便存在也还很遥远。事实上，富有戏剧性的是，就在默顿和斯科尔斯获得诺贝尔奖的第二年，这两人曾亲自出任董事会成员的著名对冲基金：美国长期资本管理公司（Long-Term Capital Management）就陷入了重大危机，被其他公司接管。而全球金融危机的爆发更是使很多人对金融世界究竟存不存在规律产生了怀疑。有人认为，早在20世纪60年代末，有“分形之父”美誉的数学家曼德布洛特

（Benoît B. Mandelbrot, 1924—）就已经提出过，金融世界在本质上是混沌的。但另一些人则认为，即便金融世界果真是混沌的，那也只不过是说我们无法进行长期预测，定量分析师仍然有可能通过分析短期规律来获取利润。究竟哪种观点正确，恐怕还有待于更多的讨论，这其中也不乏物理学家参与的余地。

在约稿信中，编辑曾建议我结合自己的经历谈谈体会，不过我想这对年轻人恐怕不会有劝导力，因为我并不是什么成功人士。我唯一能说的，是当我离开物理去做别的职业后，从未遇到过技术性的困难，所有的问题与我曾经解决或试图解决过的物理问题比起来，都显得相对简单。有时我会想到一个或许不太贴切的比喻：小时候我像很多其他小朋友一样，看电影《少林寺》入了迷，幻想着自己也能练一些武功，比如轻功。于是我让妈妈给我做了一对可以绑在腿上（但不能让小朋友们看出来）的沙袋，天天扎着走，期待有朝一日去掉绑腿后就算不能飞檐走壁，起码也能健步如飞。我觉得，学物理所受的训练就好比是扎着绑腿走路的那种锻炼，而转到别的职业后的感觉是去掉了绑腿。走路本身的难度并没有改变，但因为有了扎绑腿练就的基础，走路时就可能会觉得比较轻松。

我记得很多年前，人们曾经很看重学历，后来的一个鲜明转变是越来越多的人意识到了能力重于学历。类似地，专业曾经是很重要的求职凭据，但在日益注重能力的时代里，专业与职业的关联也在很大程度上让位给了能力与职业的关联。一个专业对口的人虽然能比其他人更快地投入工作，但这个优势往往只体现在最初的一小段时间里。一旦大家都熟悉了业务之后，究竟谁更有效率，谁更能处理复杂问题，谁更能应对尖锐挑战，终究还是要看能力。而学物理对能力的训练是比较全面的，既有严密的数学和逻辑，又能与现实数据打交道，这个专业具有广泛的

适用性是不足为奇的。

在本文的最后，请允许我再强调一次：本文的目的不是鼓励不喜爱物理的人通过学物理来达到其他目的（比方说，如果你想做的原本就是金融，那就不要去学物理），而只是想告诉喜爱物理的年轻人，学物理不是单行道，不要为出路担心，更不要因为无谓的担心而过早地放弃物理。美国物理学家费恩曼（Richard Feynman，1918—1988年）在去世前不久曾收到过一位父亲的来信，为自己即将进大学的孩子的前途问题征询意见。费恩曼在回信中提了这样一条建议：“别考虑你想成为什么，只考虑你想做什么。”

喜爱物理的年轻朋友，如果你现在想做的是学物理，那就听费恩曼的话，大胆地去做吧。

2010年4月22日写于纽约

[1]本文曾发表于《现代物理知识》2010年第3期（中国科学院高能物理研究所）。

关于普通科普与专业科普

本文的主要目的是叙述一下我对科普——尤其是数学、物理类科普——的某些零星想法，作为对拙作《黎曼猜想漫谈》的后记所提到的“普通科普”与“专业科普”这两个概念的注释，并对专业科普的价值略作评述。

我觉得普通科普（即基本不用数学公式的科普）比较适合于介绍那些容易进行通俗类比的东西，因为通俗类比是向普通读者介绍技术性内容的最有效的手段之一，通常具有将定量转化为定性，将不熟悉概念转化为熟悉概念的作用（当然，往往会因转化而导致拙作《从民间“科学家”看科普的局限性》所述的那些局限性）。对于不容易进行通俗类比的内容，普通科普则会面临不小的困难，并且常常会陷入这样的困境：即对某些无法回避的数学公式或技术性内容不得不进行缺乏类比，或类比得不太贴切的文字描述，有时甚至不得不对数学公式进行文字化的“直译”或复述——后者或许可以称为“文字公式”。

“文字公式”相较于数学公式来说，其实往往是更不容易理解的东西。事实上，从历史上讲，数学符号之所以被引入科学，乃是因为它有着文字无法替代的简单性和清晰性。从这个意义讲，从文字到数学符号乃是往简单和易于理解的方向迈出的一步，而不是相反；而对数学公式进行文字“直译”或复述，反倒是在一定程度上重新退回到了“史前”科学的繁琐、晦涩及模糊。那样的叙述虽然在表面上避免了被科普

界视为“票房毒药”的数学公式，给人以普及的印象，实际上却未必比直接使用数学公式更具普及性。因为没有数学基础的读者读到这种“文字公式”后，虽然每个字都认识，却未必能把握整句话的确切含义（或产生一个把握了的错觉）；而有一定基础的读者看了这种“文字公式”则可能会有隔膜感，会在脑子里试图将“文字公式”还原成数学公式，却远不如直接看到后者来得轻松透彻。因此，对于那种为回避数学公式而不得不诉诸“文字公式”的题材，使用“文字公式”的实际结果有可能是两头不讨好，即既不能有效地帮助普通读者理解公式的含义，也投不了有一定基础的读者所好。对于那样的题材，我觉得专业科普（即介于普通科普与专著之间、不回避数学公式的科普，有时也称为“高级科普”，不过我更倾向于“专业科普”这一术语，以避免因“高级”一词造成普通科普“低级”的不必要的攀比印象）有很大的施展余地。

当然，对数学公式与文字的难易评判不可一概而论，复杂到一定程度的数学公式自然绝非普通读者所能理解。比如黎曼-西格尔

（Riemann-Siegel）公式就是一个例子^[1]。但即便那样的公式，也有些特殊的价值，比如在向读者介绍计算黎曼 ζ 函数非平凡零点的难度时，我们固然可以搜肠刮肚地找出一系列形容词来加以描述，或者用数学家们计算零点的艰辛努力来作间接说明，但让读者亲眼看一看黎曼-西格尔公式的复杂性，也不失为是一种方法。哪怕看不懂，只当插图来看，也有可能起到一种更直接，甚至印象更深刻的说明作用。

另一方面，即便对于普通科普能够胜任的内容来说，过分排斥公式在我看来也是不必要的谨慎，甚至可以说是某种程度上的误区。对这一误区最直白的描述也许是英国物理学家霍金（Stephen Hawking）在某一版的《时间简史》（*A Brief History of Time*）的前言或后记中引述过的一句编辑的警告：每一个数学公式都会使读者减半。记得霍金在引述了

那句警告后，表示自己在书中还是用到了一个公式： $E=mc^2$ 。他并且风趣地表示，希望那不会使该书的读者减少一半。霍金的胆子算是比较大的，更多的科普作者恐怕宁肯用“能量等于质量乘以光速的平方”那样的“文字公式”来代替 $E=mc^2$ 这样的数学公式。但仔细想想，那样的“文字公式”果真比数学公式更容易普及吗？有多少读者是知道什么叫做“平方”，却不知道它在数学上是用右上角的“2”来表示的？更何况，“质量乘以光速的平方”中的“平方”究竟是指“光速”的平方，还是“质量乘以光速”的平方，在“文字公式”中是分不清的，而初中甚至高小水平的读者多半就已经知道像 $E=mc^2$ 那样的数学公式中的平方是 c 的平方，而不是 mc 的平方，因为后者会被写成 $E=(mc)^2$ ，而不是 $E=mc^2$ 。数学公式的明晰性在这么一个小小的例子中都能显现出来，普通科普却千方百计地试图避免，不能不说是某种程度上的误区。

回到黎曼猜想这一题材上来。以上所说绝不是暗示我所读过的那两本有关黎曼猜想的科普书已经陷入了那样的误区或困境^[2]。事实上，对于黎曼猜想这样一个高度技术性的数学题材来说，那两本书在深入浅出方面所做的努力是很值得钦佩的，而且它们各自都使用了少量的数学公式。不过，读者看完那两本科普后，对数学故事毫无疑问会留有印象，但对黎曼猜想本身究竟能知道多少，或许仍是可疑的。因为在对数学公式作了较大幅度的回避之后，容易出现这样的情形：即数学故事中数学的面目远比人物的面目来得模糊。而数学故事中数学的面目一旦模糊了，那么故事的背后究竟是黎曼猜想、费马猜想、还是哥德巴赫猜想，也有可能变得模糊起来。若干年之后，看过黎曼猜想、费马猜想，或哥德巴赫猜想科普书的读者或许只会记得这样的共同场景：那就是一群数学家作了各种各样的努力，经历过各种各样的趣事，试图解决一个著名的数学猜想。但他们试图解决的是什么猜想，他们各自究竟做了什

么？则有可能只是记忆中的一团迷雾。我觉得专业科普在驱散这团迷雾上也能有一定的作为，可以作为普通科普很好的补充。

以上是对《黎曼猜想漫谈》一书所采用的专业科普这一定位的一点说明。关于普通科普与专业科普这一话题本身，当然还有很多其他可以谈论的地方，绝非本文这样的零星叙述所能涵盖。

2011年3月9日写于纽约

[1]黎曼-西格尔公式是一个计算黎曼 ζ 函数非平凡零点的复杂公式，具体形式可参阅拙作《黎曼猜想漫谈》（清华大学出版社，2012年）第11章。

[2]“那两本有关黎曼猜想的科普书”指的是拙作《黎曼猜想漫谈》的后记所提到的德比希尔（John Derbyshire）的*Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics*（Joseph Henry Press, 2003）和索托伊（Marcus du Sautoy）的*The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics*（Harper, 2003）。

人名索引

A

阿克斯迪杰克 (Erik Akkersdijk)

阿罗什 (Serge Haroche)

阿西莫夫 (Isaac Asimov)

埃尔德什 (Paul Erdős)

艾弗里特 (Hugh Everett III)

艾伦伯格 (Jordan Ellenberg)

爱丁顿 (Authur Eddington)

爱因斯坦 (Albert Einstein)

安德森 (Philip Warren Anderson)

安德逊 (Carl David Anderson)

奥本海默 (Robert Oppenheimer)

B

贝尔（Jocelyn Bell Burnell）

彼得森（Ivars Peterson）

波波夫（Victor Popov）

波利策（Hugh David Politzer）

波利尼亚克（Alphonse de Polignac）

玻恩（Max Born）

伯克霍夫（George David Birkhoff）

博伊尔（Willard S. Boyle）

布莱克（Fischer Black）

布莱克特（Patrick Blackett）

布林（Sergey Brin）

布罗特（Robert Brout）

C

查基尔（Don Zagier）

陈景润

茨威格（George Zweig）

D

达文波特（Harold Davenport）

戴维森（Morley Davidson）

丹尼特（Daniel C. Dennett）

德比希尔（John Derbyshire）

德布朗基（Louisde Branges）

德斯里奇（John Dethridge）

丁格尔（Herbert Dingle）

E

厄斯特勒（Joseph Oesterlé）

F

法尔廷斯（Gerd Faltings）

法捷耶夫（Ludvig Faddeev）

凡尔纳（Jules Verne）

范斯托克姆（Willem Jacob van Stockum）

弗兰克林（Philip Franklin）

G

盖尔曼 (Murray Gell-Mann)

盖姆 (Andre Geim)

高登 (Walter Gordon)

高锟 (Charles K. Kao)

高特 (John Richard Gott III)

戈德菲尔德 (Dorian Goldfeld)

戈德斯通 (Daniel Goldston)

哥德尔 (Kurt Gödel)

格拉肖 (Sheldon Lee Glashow)

格兰维尔 (Andrew Granville)

格雷克 (James Gleick)

格林 (Brian Greene)

格林伯格 (Oscar W. Greenberg)

格娄斯 (David Gross)

格罗滕迪克 (Alexander Grothendieck)

H

哈代 (Godfrey Hardy)

海森伯 (Werner Heisenberg)

韩武永 (Moo-Young Han)

赫克斯利 (Martin Huxley)

赫兹 (Heinrich Hertz)

怀尔斯 (Andrew Wiles)

惠勒 (John Archibald Wheeler)

霍夫施塔特 (Douglas R. Hofstadter)

霍金 (Stephen Hawking)

霍克汉姆 (George Hockham)

J

季米特洛夫 (Vesselin Dimitrov)

嘉当 (Élie Cartan)

伽利略 (Galileo Galilei)

伽莫夫 (George Gamow)

金明迥 (Minhyong Kim)

K

卡西米尔 (Hendrik Casimir)

康拉德 (Brian Conrad)

考克 (Bruce Cork)

科尔曼 (Sidney Coleman)

科先巴 (Herbert Kociemba)

克莱因 (Oskar Klein)

孔克拉 (Dan Kunkle)

库伯曼 (Gene Cooperman)

L

拉杜 (Silviu Radu)

拉普拉斯 (Pierre Simon Laplace)

莱特兄弟 (Wright brothers)

兰金 (Robert Alexander Rankin)

朗道 (Lev Landau)

朗之万 (Paul Langevin)

劳厄 (Max von Laue)

勒纳 (Philipp Lenard)

勒维耶 (Urbain Le Verrier)

雷登弗罗斯特 (Johann Gottlob Leidenfrost)

李特伍德 (John Littlewood)

李政道⁹⁶

里德 (Michael Reid)

里奇 (Giovanni Ricci)

梁灿彬

鲁比克 (Ernö Rubik)

伦德勒 (Wolfgang Rindler)

罗基奇 (Tomas Rokicki)

洛伦兹 (Edward Norton Lorenz)

洛伦兹 (Hendrik Lorentz)

M

马赫 (Ernst Mach)

麦瑟尔 (David Masser)

梅尔 (Helmut Maier)

梅里利斯 (Philip Merilees)

米尔斯 (Robert Mills)

米斯纳 (Charles W. Misner)

米歇尔 (John Michell)

密立根 (Robert Andrews Millikan)

闵科夫斯基 (Hermann Minkowski)

莫勒 (Christian Møller)

莫里斯 (Mike Morris)

默顿 (Robert C. Merton)

N

南部阳一郎 (Yoichiro Nambu)

内曼 (Yuval Ne'eman)

牛顿 (Isaac Newton)

诺维科夫 (Igor Novikov)

诺沃肖洛夫 (Konstantin Novoselov)

O

欧几里得 (Euclid)

欧勒特 (Walter Oelert)

P

庞加莱 (Henri Poincaré)

泡利 (Wolfgang Pauli)

培根 (Francis Bacon)

佩雷尔曼 (Grigory Perelman)

佩奇 (Larry Page)

蓬皮埃利 (Enrico Bombieri)

平兹 (János Pintz)

Q

齐奥尔科夫斯基 (Konstantin Tsiolkovsky) 156, 174, 178, 180-183

钱德拉塞卡 (Subrahmanyan Chandrasekhar) 74

乔伊斯 (James Joyce)

S

萨根 (Carl Sagan)

萨哈洛夫 (Andrei Sakharov)

萨克斯 (Rainer Sachs)

萨拉姆 (Abdus Salam)

塞克斯尔 (Roman U. Sexl)

赛格雷 (Emilio G. Segrè)

桑德拉拉扬 (Kannan Soundararajan)

施皮罗 (Lucien Szpiro)

施瓦西 (Karl Schwarzschild)

史密斯 (George E. Smith)

舒斯特 (Arthur Schuster)

斯科尔斯 (Myron Scholes)

索恩 (Kip S. Thorne)

索末菲 (Arnold Sommerfeld)

索托伊 (Marcus du Sautoy)

索兹曼 (Barry Saltzman)

T

塔姆 (Igor Tamm)

特勒 (Edward Teller)

汤姆逊 (Joseph John Thomson)

陶哲轩 (Terence Tao)

特·胡夫特 (Gerard 't Hooft)

梯普勒 (Frank J. Tipler)

托雷提 (Roberto Torretti)

托曼 (Richard C. Tolman)

W

瓦法 (Cumrun Vafa)

外尔 (Hermann Weyl)

望月新一 (Shinichi Mochizuki)

威顿 (Edward Witten)

威尔斯 (H. G. Wells)

威廉森 (Jack Williamson)

韦尔切克 (Frank Wilczek)

韦斯科夫 (Victor Weisskopf)

维因兰德 (David Wineland)

温伯格 (Steven Weinberg)

文卡塔斯 (Akshay Venkatesh)

沃尔德 (Robert M. Wald)

沃勒 (Ivar Waller)

X

西斯尔斯韦特 (Morwen Thistlethwaite)

希尔伯特 (David Hilbert)

希格斯 (Peter Higgs)

休伊什 (Anthony Hewish)

薛定谔 (Erwin Schrödinger)

Y

亚伯拉罕 (Max Abraham)

亚当斯 (John Couch Adams)

杨振宁¹²⁴

伊尔迪里姆 (Cem Yildirim)

英格勒特 (François Englert)

约纳-拉西尼奥 (Giovanni Jona-Lasinio)

Z

张伯伦 (Owen Chamberlain)

张益唐

术语索引

ABC@Home

ABC猜想

CCD

Excite

Math Overflow

SETI

$SU(2)$

$SU(3)$

TNT

$U(1)$

ZetaGrid

A

爱因斯坦-嘉当理论

爱因斯坦升降机

暗能量

暗物质

B

白矮星

保罗阱

贝林番特张量

闭合类时曲线

边带冷却技术

表兄弟素数

波函数

波利尼亚克猜想

玻色子

不确定原理

布莱克-斯科尔斯模型

C

参照系

测地线

超弦理论

陈氏定理

虫洞

虫洞时间机器216, 221

磁矩81

D

大气伽马切伦科夫成像望远镜

大型强子对撞机

等效原理

狄拉克方程

第二哈代-李特伍德猜想

第二宇宙速度

第三宇宙速度

第一哈代-李特伍德猜想

第一宇宙速度

点粒子

电磁观

电弱统一理论

电子简并压

定量分析师

定量金融

丢番图分析

动力气象学

对称性

对称性自发破缺

多世界诠释220

E

二体问题

F

反常

反粒子

反物质

反重子

仿射联络

非线性

菲尔茨奖

费恩曼图

费马猜想

费马大定理

费米伽马射线太空望远镜

费米子

分布式计算

负能量

富勒烯

伽马射线暴

伽马射线耀斑

G

概率启发式理由

戈德斯通定理

戈德斯通粒子

哥德巴赫猜想

哥德尔宇宙

格点量子色动力学

谷歌

谷歌矩阵38

惯性参照系

光电效应

光纤

广义黎曼猜想8

广义相对论

规范对称性

规范理论

H

哈勃太空望远镜

哈代-李特伍德猜想

褐矮星

黑洞

蝴蝶效应

互素

火箭

霍金辐射

J

机械观

几何动力学

简并

渐近自由

胶子

金融衍生品

近周期性

经典电子论

晶体管

决定论

K

康普顿波长

科尔曼-温伯格机制

科普

可穿越虫洞

可见物质

克莱因佯谬

克隆

空穴理论

夸克

夸克星

L

拉氏量

劳厄定理

雷登弗罗斯特效应

类时曲线

类星体

离子阱

黎曼-嘉当几何

黎曼猜想

黎曼几何

黎曼空间

里德堡原子

理想时钟

量子场论

量子电动力学

量子霍尔效应155

量子力学

量子色动力学

量子引力理论

量子涨落

鲁比克方块

旅行者一号

孪生素数

孪生素数猜想

孪生素数常数

裸质量

洛伦兹对称性

洛伦兹方程组

洛伦兹群

M

马尔可夫过程

马尔可夫链

马尔可夫链基本定理

马西森-帕帕佩特鲁-狄克逊方程

民间“科学家”225-232， 244

民科

闵科夫斯基度规

闵科夫斯基空间

魔方

N

纳米管

挠率

能量动量张量

诺特定理

诺维科夫自洽性假设218

O

耦合常数

P

庞加莱猜想

庞加莱群

庞加莱张力

泡利不相容原理

佩奇排序

彭宁阱

平行宇宙

普朗克长度

普朗特常数

Q

齐奥尔科夫斯基公式

奇怪吸引子

强孪生素数猜想

曲率

曲率张量

曲速引擎

圈量子引力

圈图

全内反射

群论

“钱德拉”X射线太空望远镜

R

热力学第二定律

日本学术奖章

S

萨哈洛夫条件

三体问题47

色荷

筛法

上帝之数

神舟五号

生命传输机

施瓦西解

石墨

石墨烯

时间机器

时空泡沫

时序保护假设

时钟假设

时钟延缓

时钟佯谬

世界线

视界204

手征对称性

手征凝聚

手征外推

手征微扰理论

数论

双生子佯谬

双重狭义相对论

四色猜想

四色定理

素矩阵

素数

素数定理

索性修正

随动惯性系

随机矩阵

随机性修正

T

太空电梯

太空望远镜

汤川耦合

天气学

W

微型黑洞

卫星一号

X

西斯尔斯韦特算法

希格斯场

希格斯机制

希格斯粒子

系综

狭义相对论

先驱者10号

先驱者11号

相对论

相对性原理

香蕉皮机制

星际旅行

悬挂网页

薛定谔的猫

Y

雅虎

湮灭

贗戈德斯通粒子

杨-米尔斯理论

以太

因果律

因果佯谬

永动机

宇称

宇宙间虫洞

宇宙内虫洞

宇宙微波背景辐射

宇宙学

原子钟

圆法

云室

Z

张力

真空期待值

真空色散

真空态

正电子

正矩阵

指标函数

质量隙

中子星

终极理论

重整化

重子

重子数

周期性

专业科普

转移矩阵

准粒子

自能

自旋

理解科学丛书·赵峥科普著作

THE QUEST OF
THE TRUTH
from A.Einstein to S.W.Hawking

物含妙理总堪寻

从爱因斯坦到霍金

赵峥◎著

讲解20世纪以来一个个激动人心的
物理学发现及其基本原理，
勾画现代人应该具备的科学文化视野。

中国大学视频公开课TOP10人气榜中唯一的一门自然科学课程，
一部深入浅出、引人入胜的近代物理学“教科书”。

清华大学出版社

目 录

[前言](#)

[第一讲 爱因斯坦与物理学的革命](#)

[第一讲附录 狭义相对论的创立](#)

[第二讲 弯曲的时空——广义相对论](#)

[第二讲附录 爱因斯坦与广义相对论](#)

[第三讲 白矮星、中子星与黑洞](#)

[第三讲附录 漫谈黑洞（I）](#)

[第四讲 霍金与黑洞](#)

[第四讲附录 漫谈黑洞（II）](#)

[第五讲 膨胀的宇宙](#)

[第六讲 时空隧道与时间机器](#)

[第七讲 激动人心的量子物理](#)

[第七讲附录 玻尔对爱因斯坦光子箱实验的答复](#)

[第八讲 比一千个太阳还亮](#)

[第八讲附录 汤川对介子质量的估计](#)

[第九讲 漫步太阳系](#)

[第十讲 时间之谜](#)

[插页诗句的注释与随想](#)

[主要参考书目](#)

[后记](#)

[返回总目录](#)

作者简介

赵峥 北京师范大学物理系教授，博士生导师。1967年毕业于中国科技大学物理系。1981年于北京师范大学天文系获硕士学位。1987年于布鲁塞尔自由大学国际理论化学与理论物理研究所获博士学位。曾任北京师范大学研究生院副院长，物理系主任，中国引力与相对论天体物理学会理事长。长期从事理论物理的教学与研究。在相对论与黑洞领域发表科研论文100余篇。曾讲授广义相对论、量子力学、统计物理、黑洞物理、量子场论等课程。同时参加教学研究、科普教学。开设“从爱因斯坦到霍金的宇宙”等公选课，并在许多院校做科普报告。主要著作有《黑洞的热性质与时空奇异性》（获第十二届中国图书奖）、《广义相对论》、《广义相对论基础》、《黑洞与时间的性质》、《黑洞与弯曲的时空》。以及科普类图书《探求上帝的秘密》（获第十一届中国图书奖）、《物理学与人类文明十六讲》、《相对论百问》等。



谁家吹笛画楼中，
断续声随断续风。
响遏行云横碧落，
清和冷月到帘栊。

唐·赵嘏

处处中秋此月明，
天涯何处亦群英。
须怜绝学经千载，
莫负男儿过一生。

明·王阳明



绘画：张京

The Quest of the Truth: From A.Einstein to S.W.Hawking

物含妙理总堪寻
——从爱因斯坦到霍金

赵 峥 编著

清华大学出版社
北 京

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

物含妙理总堪寻：从爱因斯坦到霍金 / 赵峥著. 北京：清华大学出版社，2013

（理解科学丛书）

ISBN 978-7-302-34049-2

I. ①物... II. ①赵... III. ①物理学—青年读物②物理学—少年读物 IV. ①04-49

中国版本图书馆CIP数据核字（2013）第237625号

责任编辑：朱红莲

封面设计：蔡小波

责任校对：赵丽敏

责任印制：王静怡

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社总机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市金元印装有限公司

经 销：全国新华书店

开 本：165mm×240mm 印 张：18 插 页：1 字 数：
275千字

版 次：2013年12月第1版 印 次：2013年12月第1次印刷

产品编号：055298-01

前言

作者长期以来在北京师范大学开设科普讲座“从爱因斯坦到霍金的宇宙”，历时20余载，并在一些院校和单位举办过不同形式的讲座和公开课，重点介绍物理学和天文学领域的科普知识、科研前沿，以及科学发现的曲折历程，目的在于扩展学生的科学视野，增强学生的创新能力。

本书总结历次讲座和公开课的核心，以演讲集的方式呈现给读者，内容主要包含：爱因斯坦与相对论、弯曲的时空、黑洞、宇宙的演化、量子论的创建与争论、原子弹与核能的和平利用、天文学的若干知识、对时间本质的探索等等。内中涉及一般读者感兴趣的双生子佯谬、宇宙创生、时空隧道、时间机器、薛定谔猫、量子力学的多次论战、黑洞的神奇性质等问题。

在演讲集中作者力图把科学家们作为有血有肉的人展现在大家面前，通过科学家千姿百态的人生经历和科学发现“山重水复”、“柳暗花明”的历程，尽可能使读者看到真实的历史和鲜活的人物形象，从而了解到科学家不一定是完人，但都是创造历史的伟人。

当前中国正处在大发展、大变革的时代，年青人有着施展才华的无限机遇，也面临着各种无法预料的风险和挑战。

曾子勉励过年轻人：

士不可以不弘毅，
任重而道远。

清代诗人赵翼也说：

江山代有才人出，
各领风骚数百年。

本书书名源于乾隆的一副对联：

境自远尘皆入咏，
物含妙理总堪寻。

这副对联位于颐和园万寿山，铜亭附近的一座石碑坊上。

赵 崢
2013年初秋于北京

第一讲 爱因斯坦与物理学的革命

我们现在开始讲第一讲，《爱因斯坦与物理学的革命》，就是简单介绍爱因斯坦在相对论和量子论建立时的贡献。

请大家看一下图1-1这张照片，这张照片跟大家通常看到的爱因斯坦不太一样。通常看到的都是那个头发乱糟糟、满脸皱纹、叼个烟斗的爱因斯坦。大家都觉得，哎呀，这个脑袋聪明得不得了！其实那个脑袋已经不太行了，行的是什么呢？行的是图1-1中的脑袋，是他发表狭义相对论的时候、26岁左右的脑袋。



图1-1 青年爱因斯坦

我觉得现在有很多宣传给年轻人造成一个印象——重大成就都是老头老太太发现的，其实不完全是这样。一般来说，做出重大发现的以中青年人居多，很多还是青年人，他们在二十多岁、三十多岁时就做出了重大贡献。到了四五十岁以后，基本都是学问大了，但是创新性的贡献不太大了，年老了以后，奇思异想少了，闯劲小了，人的创新能力也就大大下降了。所以同学们要努力，要争取在中青年时代做出成就。

1. 量子论的诞生

两朵乌云

好，我们现在就讲一下20世纪初的这次物理学革命。当时有一件很有意义的事情，就是1900年的4月27日，英国皇家学会为迎接新世纪的来临，开了一次庆祝会。在这个会上，德高望重的物理学权威开尔文勋爵发表了一个很著名的演说，这个演说中说“物理学的大厦已经建成，未来的物理学家只需要做些修修补补的工作就行了”。这是因为那时牛顿的力学已经完美地建立起来，随后发展成拉格朗日的分析力学；牛顿的光学也发展起来了，后来又被波动光学所取代；电磁学也发展起来了；热学也已经发展起来了。所以物理学家们充满了信心，认为物理学已经基本完成任务了。

但是另一方面，开尔文还有一双慧眼。他说，现在还存在两个问题。而且他认为这两个问题比较重要。于是他接着说：“现在明朗的天空还有两朵乌云：一朵与黑体辐射有关，另一朵与迈克耳孙实验有关。”直到现在我们仍经常谈起“两朵乌云”，因为开尔文这些话太有名了。在开尔文讲了这段话不久，就从这两朵乌云里面诞生了量子论和相对论。

当年的年底就从第一朵乌云中诞生了量子论，是由普朗克提出来的。五年之后从另一朵乌云中诞生了相对论，是由爱因斯坦提出来的。

而且爱因斯坦在那一年把普朗克的量子论发展成光量子理论，也就是今天的光量子理论。所以开尔文说的“两朵乌云”非常有名。今天我们还可以看到一些物理学上的困难，不断地有人说这又是一朵乌云，那又是一朵乌云。其实全都不灵，说的都不对，说明这些预言的人水平不够。

黑体辐射之谜

我们现在来看看第一朵乌云。第一朵乌云是黑体辐射。1870年，普法战争法国战败。法国战败以后，支付给普鲁士一大笔战争赔款，并且把阿尔萨斯和洛林两个省割让给普鲁士。这件事情大家在《最后一课》里可以读到。这两个省对普鲁士至关重要，因为这两个省靠着普鲁士的鲁尔区，鲁尔区产煤，没有铁；而法国这两个省有铁，没有煤，现在都归了普鲁士。同时普鲁士又得到了一大笔战争赔款。当时普鲁士的统治集团还是有所作为的，想把他们的国家搞得富强起来。他们就用这笔钱来发展钢铁工业，力图建立德意志帝国，把普鲁士从一个以生产土豆为主的国家变成一个以生产钢铁为主的国家。我们现在也在经历这样一个阶段，从一个农业国走向一个工业国，实际上这是一个伟大的进步。但是，炼钢需要控制炉温，炉温怎么控制呢？你不能塞一个温度计进去，那一下就烧化了。怎么办呢？就在高炉上开一个小孔，看它射出来的热辐射，根据这种热辐射在不同波长的能量密度分布，可以得到一些实验点，就是图1-2上这一个个的圆圈点。将这些圆圈点连起来可以形成一条实验曲线，根据这条实验曲线就可以判定炉温。比较著名的是维恩位移律，这个定律指出，热辐射的能量密度取极大值处的波长，也就是实验曲线的最高点处的波长 λ_m 与温度的乘积是一个常数，用这个式子可以很容易地定出炉温。这种热辐射叫做黑体辐射。

$$T\lambda_m = b \quad (1.1)$$

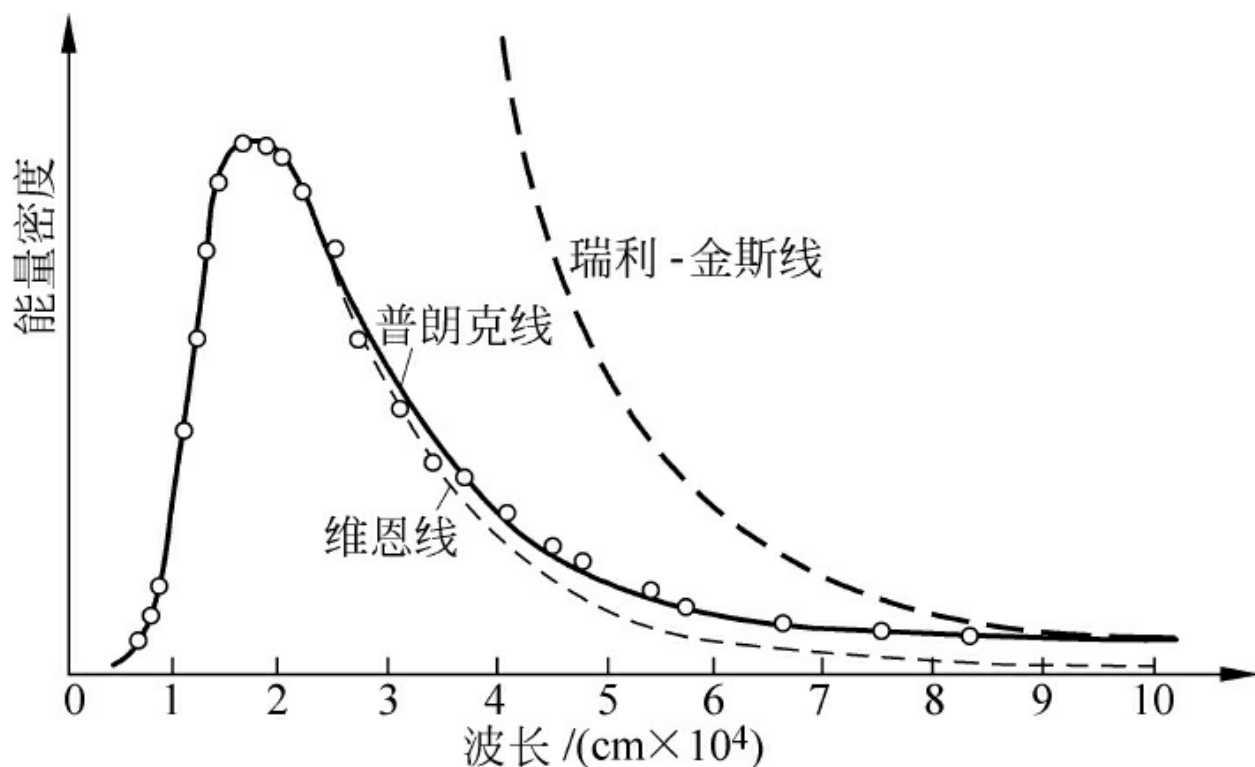


图1-2 黑体辐射

为什么黑体辐射会表现出这样一条曲线呢？当时物理学家们都搞不清楚。那个时候原子论还没有被大家普遍接受。当时在解释这种黑体辐射的时候，认为每一个原子都像一个谐振子，不承认原子论的人也可以从别的角度来看这个辐射源，反正它都像一个一个的谐振子。它吸收辐射，振动就加剧；它放出辐射，振动就变缓慢。

当时英国正在开展工业革命，也在发展钢铁工业。英国的瑞利和金斯根据这样的一种物理构想，得到了一条曲线。这条曲线在长波波段与实验点符合得很好，但短波波段是无穷大，这就是著名的紫外光灾难。德国的维恩使用的模型跟他们的模型不大一样，但也得到了一条曲线，它在短波波段与实验点符合得不错，而在长波波段偏离了实验点。这就是当年开尔文谈到的黑体辐射困难。不过开尔文的原话实际上不是谈黑体辐射，而是谈固体比热。但固体比热问题大家一般不大熟悉，也不大直观。说到黑体辐射时，你可以简单地告诉别人是怎么回事。实际上黑

体辐射和固体比热说的是同一个困难。

普朗克的突破——量子假设

那时德国的理论物理学家普朗克，也研究这个问题，但始终不能得到一个很好的结果。有一次，他偶然发现，假如认为谐振子放出辐射和吸收辐射是一份一份的，不是连续的，那么就可以得到一条曲线，这条曲线跟实验点很好地相符。但是辐射怎么可能是一份一份的呢？当时已经知道热辐射与光辐射本质相同，它们都是电磁波，都是连续的，怎么可能变成“一份一份”不连续的呢！所以他对自己的这个发现很犹豫，一方面觉得很惊喜，另一方面也很担心。因为他已经是教授了，万一闹个笑话就不大好。

有一次他在学校里给学生作报告介绍自己的这个发现。他讲得非常保守，以至于有一些学生听完了以后觉得今天白来了一趟，普朗克教授什么也没有讲出来。但是，在跟他儿子出去散步的时候，普朗克说：“你爹我呀，现在做出了一个发现，这个发现如果被证明是正确的，将可以跟牛顿的成就相媲美。”可见他对这个发现是很重视的。因为物理学是一门实验的科学，测量的科学，你的理论再好，如果不能跟实验相符，你的理论肯定被否定。反过来，你的理论让人感到非常牵强，但是能够解释实验，大家就可以接受。所以大家带着很大的怀疑接受了普朗克的这个理论。物理学家们普遍觉得他这个理论尽管可疑，但跟实验一致，还可以勉强接受。

当时，普朗克是这么认为的：“热辐射从原子中射出来的时候是一份一份的，吸收的时候也是一份一份的，但是辐射脱离原子之后，在空间中传播的时候还是连续的，不是一份一份的。”他这样解释自己的观点，但大家都听不懂。有一个记者就问他，说：“普朗克教授，您一会儿说辐射是连续的，一会儿又说它是不连续的，那么它到底是连续的还是不连续的？”普朗克说：“有一个湖，湖里头有很多的水，旁边有一个水缸，里头也有水，有人用小碗把缸里的水一碗一碗地舀到湖里，你说

这水是连续的还是不连续的？”我认为这个解答清楚地阐明了他对这个问题的看法。

争论：量子还是光量子？

五年之后，德国的《物理年鉴》，相当于中国的《物理学报》，收到了一个年轻人的论文，是解释光电效应的。这个年轻人叫爱因斯坦，当时大家完全没有听说过他。这篇文章说：光辐射在脱离原子以后依然是一份一份的。普朗克一看，不同意这个观点，但是这个理论能够解释光电效应。普朗克表现出大家风范，一方面同意发表这篇论文，另一方面写信给爱因斯坦，还很虚心地向他请教，问他：你这是怎么回事啊？

爱因斯坦当时是个无名小卒，拿到普朗克这封信的时候，他都不敢相信这真是大物理学家普朗克给他写的。他一想：这准是他那几个朋友，“那几个小丑”在捣蛋，跟他开玩笑，冒充普朗克给他写了这封信。此时他的夫人正在那里洗衣服，一把就把那封信抢过来，一看，说：“这封信是从柏林寄出来的。”而他们当时住在瑞士，她说：“他们不可能到柏林去给你发这封信，来给你捣蛋啊。”爱因斯坦仔细一看，真是普朗克的！

后来普朗克还派他的助手劳埃来拜访爱因斯坦，跟爱因斯坦讨论这个问题。普朗克一直认为爱因斯坦对量子的解释是不对的。爱因斯坦随后又连续写了几篇论文，包括相对论的论文，都是普朗克审的，普朗克都同意发表了，而且都高度赞扬。只有这一篇论文，普朗克非常有保留。普朗克在给维恩的信里就讲：“当然了，爱因斯坦的这个观点肯定是错误的。”但是他还是支持爱因斯坦这篇论文的发表。直到1913年，普朗克推荐爱因斯坦担任德国普鲁士科学院院士的时候，他为其写的推荐信里还在讲，说爱因斯坦做出了很多伟大的成就，等等，之后，他说：当然了，我们也不能对一个年轻人有太多的苛求，我们还是应该允许他有一些错误。比如他对光量子的解释好像就是不大对的，但是，这丝毫掩盖不了他的光辉……说了这么一段。

接着没有过几年，诺贝尔奖评委会开始评奖，大家都认为应该给爱因斯坦发奖，理由是什么？有很多人认为是相对论，但有一些人说相对论根本看不懂啊，万一是错的怎么办呢？于是大家讨论了半天，最后达成一个妥协，以爱因斯坦解释光电效应和在物理学其他方面的成就授予他诺贝尔物理学奖，就没提相对论。而且评委会的秘书在给爱因斯坦写信通知他获奖时还写道：“当然了，这次给你授奖，没有考虑你在相对论（即狭义相对论）和引力论（即广义相对论）方面所作出的贡献。”就是说没有因为发现相对论给他授奖。也可能有一些人还准备给他第二次授奖，但后来诺贝尔奖评委会不想给一个人颁两次奖。诺贝尔科学奖真正得过两次的只有两个人，一个是居里夫人，另一个是巴丁，其他人都没有。

2. 爱因斯坦的成长历程

家里来的大学生

好，我们现在来看看这个爱因斯坦是怎样一个人。他是个犹太人，父母都很喜欢音乐。他父亲开了个小工厂，大概就是几百个工人，一个小企业家。他的堂叔是这个厂的工程师。看来是个家庭企业，在慕尼黑经营。爱因斯坦出生不久，他们家就搬到了慕尼黑。爱因斯坦小时候说话很晚，一直到三岁的时候才能跟人讲得比较明白，所以大人都觉得这孩子是不是智力有问题啊！小爱因斯坦也不大注意大人们在谈论什么，他就在那儿摆弄自己的东西。有一次他父亲给他带来一个罗盘，他高兴得不得了，就成天摆弄那个罗盘。他基本不注意别人在干什么，他提的问题常常跟大人们正在谈论的东西没有关系，而是他自己在想的东西。不过，小爱因斯坦喜欢看课外书。当时德国的犹太家庭有一个习惯，中产阶级以上的犹太家庭一般都会在周末的时候，接待一位贫穷的犹太大学生到自己家度周末。他们家也来了一个，是一位医学院的学生。这学生来了以后，爱因斯坦很喜欢他，虽然跟父母谈话不多，但是他跟这个

小伙子谈话很多。这个小伙子发现爱因斯坦爱看书，就把各种各样的书都带来给爱因斯坦看，科普的、数学的，甚至哲学的。他很高兴，翻看了很多书，也不知道看懂看不懂，反正都在那儿很专心致志地看。所以他的知识很丰富。看来这个大学生的出现，对爱因斯坦的智力启蒙产生了作用。

不受学校欢迎的学生

小爱因斯坦在学校里是不大受欢迎的，有几个原因，其中之一是他的功课一般。对此老师倒不会对他有什么想法，但他还有两个“短处”：一他是犹太人，德国那时有种族歧视，由于犹太人有钱，因此对犹太人是既看不起又羡慕；二他是无神论者，不相信上帝，这在当时是个严重的问题。所以学校觉得这个孩子比较烦人。另外，小爱因斯坦看的课外书多，又爱乱想，净问一些老师答不出来的问题，老师觉得很丢面子。当时德国是军国主义教育，老师都是居高临下地对待学生。老师好像什么都懂：“啊，这个你还不会！”结果小爱因斯坦问的问题老师不会。于是老师觉得很下不来台，就比较烦他。

小爱因斯坦上中学以后，他们家买卖做得不行，全家迁往意大利，投奔爱因斯坦家族的亲友，只把小爱因斯坦一个人留在慕尼黑，安排他进入一所重点中学学习。在那里老师们仍然不喜欢他，觉得这个小犹太人功课一般，不相信上帝，还总问老师答不出来的问题，有损老师的面子和学校的声誉。最后爱因斯坦在学校里感觉压力太大，待不下去了，于是他就找到自己所在街区的那个经常给他们家看病的社区医生，开了一份患神经衰弱的证明，准备休学半年去纾解一下压力。可他的证明还没拿出来呢，老师就跟他说校长找他。校长一见面就劝他退学，他一听退学，吓了一跳。这怎么跟父母交代啊！后来一想，也好，以后就再也不用来这所学校了。于是他愉快地接受了校长的建议，退学去意大利投奔自己的父母。

阿劳中学——孕育相对论的土壤

在意大利待了一段时间以后，他还是想上大学。爱因斯坦的父亲希望他回德国上，因为他的母语是德语，而且德国的科学技术比意大利先进。但他特别讨厌德国的教育方式，不愿意去。他父亲最后同意了，并建议他去瑞士，瑞士有德语区和法语区。德语区跟德国一样讲德语。于是他就去投考了苏黎世工业大学的师范系，这是一个培养大学和中学数学、物理老师的系。第一年没有考上，没考上的原因之一是他中学课程没有学完，当然他功课也一般。

爱因斯坦只好准备第二年再考，于是他在瑞士的阿劳州立中学上了一年补习班。爱因斯坦一生对学校都没有好印象，他认为学校的教育都过于呆板，把学生的思想都给束缚死了。他后来回忆说：“我很幸运，我属于少数没有被束缚死的人之一。”爱因斯坦唯独赞扬的就是他上补习班的阿劳中学。瑞士的中学跟德国的中学风格非常不一样，给学生充分的自由，学习上的自由、生活上的自由，老师非常平等地与学生进行讨论。

所以爱因斯坦没有任何压力，度过了愉快的一年，而且思考了一些问题，包括最早引导他走向相对论的那个追光悖论。这个思想实验就是那时候产生的，因为那时他有闲工夫去乱想。要是学习压力太大，学生根本没有时间去思考。但爱因斯坦在阿劳中学有充分的可以自由支配的时间去遐想，他经常想入非非。当时人们已经认识到光是电磁波。有一次他想，假如一个人追上光，跟光一起跑，能看到什么呢？大概能看到一个不随时间变化的波场。可谁也没见过这种状况，这是怎么回事呢？这个思想实验使他认识到光相对于任何人都是运动的，不可能静止。这个思想实验伴随了他十年，最后把他引向相对论的创建。

不平常的大学生涯

上了一年补习班后，爱因斯坦考上了苏黎世工业大学师范系。那时他非常高兴，他很喜欢物理，但听课课后却大失所望。讲课的物理教授是韦伯，这个韦伯不是命名为磁学单位的那个韦伯，而是个电工专家。他

讲的物理全都是跟实际联系非常密切的，他不大重视理论。可是爱因斯坦对电工不感兴趣，他感兴趣的是比较深的理论问题。爱因斯坦问老师的一些理论问题，韦伯也不会，所以他对韦伯讲的课没有兴趣。韦伯也对他印象不大好，觉得爱因斯坦不但不来听课，而且一点礼貌都没有：不叫我“韦伯教授”，居然叫我“韦伯先生”。那个“先生”估计是“Mr”那种称呼，不是特别尊敬的男士之间的称呼。

教授职位在德国是非常难得的，德国、英国以至整个欧洲，一个系通常就一个教授，这种体制绝对能够保证教授的质量。当然也有弊病，老先生不死，年轻人没法子，升不上去啊！

在那种情况之下，爱因斯坦就不去听韦伯的课了。教数学的是闵可夫斯基。现在理工科的学生在相对论中都看到过闵可夫斯基这个名字，相对论中用到了闵可夫斯基时空。这位闵可夫斯基小时候是个神童。他们弟兄几个都非常聪明。聪明到什么程度呢？在上小学时，他们与那位大数学家希尔伯特是同学。他们聪明到让希尔伯特对自己都没有信心了，回家跟父母讲：我可能不行，他们那哥几个才真聪明呢！结果呢，后来闵可夫斯基兄弟没什么太大的成就，希尔伯特反而成为数学大师。过去一百多年中两个最杰出的数学大师，一个是希尔伯特，另一个是法国的庞加莱。而这位闵可夫斯基还是靠着他的学生爱因斯坦最后出名的。当然，他后来研究爱因斯坦的相对论也有贡献。可见神童不神童不是最重要的。

爱因斯坦不去听课，每天躲在他租的一个小阁楼里。因为国外的大学通常不提供那么多宿舍，学生大都是在校外租当地居民的房子，学校附近的居民也靠出租房屋作为家庭收入的一部分。

就这样，爱因斯坦租了一个小阁楼，买了一些当时德国的著名物理学家的著作，比如赫兹啊、赫姆霍兹啊这些人的，他每天躲在小阁楼里看书。他也不是完全不去学校，一般是下午五点放学后他就去了。去干嘛呢？两件事情，一件事情就是跟同学们到咖啡馆喝咖啡，讨论讨论，

问问：“你们课堂上听了些什么啊？”同时告诉他们自己看了些什么书，交流交流。另外一个就是到实验室做实验。德国大学和瑞士大学的实验室都是开放的。瑞士在这点上跟德国是很相近的，学生可以随时进来做实验。咱们国家现在还没做到这一点，我想咱们国家要想成为一个创新型的大国，大学应该做到让学生能够进实验室自主地做实验。

米列娃与格罗斯曼

那么，爱因斯坦不去听课有没有问题呢？有。因为他需要有人帮他记笔记。不过没有关系，他们班唯一的一个女生米列娃跟他关系很好，也爱听他神侃，愿意帮他记笔记。但是米列娃功课一般，到了考试的时候，单靠米列娃的笔记不行。不过他们班还有一个优秀的学生叫格罗斯曼，这是一位标准的好学生，每天西服革履，领带打得非常好，皮鞋锃亮，功课又好，对老师又有礼貌，字也写得漂亮，是爱因斯坦的好朋友。爱因斯坦考试前几个星期就跟他借笔记，他都慷慨地借给爱因斯坦。我们当过学生的都知道，考试后借笔记一般问题不大，考试前借笔记那我还得看呢，所以格罗斯曼还真是不错，每次都借给爱因斯坦。爱因斯坦拿到这个笔记，突击两个星期，然后就去参加考试，一考还就考过去了。考过去之后，他就跟别人发表感想：这门课简直一点意思都没有。你想，这种学习方式他能感到有意思吗？肯定感觉没有意思。但他还是学到了很多，他主要通过自学学到了很多，就这样直到该毕业了。

生活的辛酸

毕业的时候，格罗斯曼和另外一个同学被闵可夫斯基留下来当数学助教，爱因斯坦想韦伯大概会把他留下来当物理助教了，结果韦伯不要他，也没要他们班的其他几个学生，而是从工科系留了两个同学。

爱因斯坦一时找不到工作，非常狼狈。曾有一个同学帮他找了一份在另外一座城市的、三个月的中学代课老师的工作，爱因斯坦写了一封感谢信，我看过那封信，简直是感激涕零啊！可见他当时真是很困难

了。他在电线杆上贴广告，说他可以教数学、物理、小提琴，一个小时多少钱，也没什么人找他。

爱因斯坦当时倒霉的事不止是找不到工作，婚姻也成了问题。因为他与米列娃结婚的事遭到他父母的坚决反对。为什么呢？米列娃出身“不好”，不是犹太人，而是属于被压迫民族的塞尔维亚人。再有呢，米列娃有残疾，她腿瘸，先天性的有的一些什么病，几本书上写的病不大一样，反正就是瘸得比较厉害。爱因斯坦的父母觉得这个女孩子怎么能配上我们的儿子呢！非常不满意。但是爱因斯坦呢，父母越不满意他越要跟米列娃好，于是就处于一种僵持状态。婚姻上碰壁，工作也解决不了。

3. 爱因斯坦的奇迹年

时来运转

直到1902年，终于时来运转了。先是在父亲临终时，爱因斯坦回意大利去看他。他父亲还是很喜欢自己的孩子，既然儿子这么坚持，就算了吧，同意了这门婚事。犹太人跟我们中国过去传统的家庭差不多，父亲是家长，父亲同意了，母亲不同意也没办法。所以，在他母亲很不情愿的情况之下，爱因斯坦被获准跟米列娃结婚。

爱因斯坦要结婚了，可是没有钱。这个时候格罗斯曼出面帮爱因斯坦找了份工作。格罗斯曼的父亲有一个朋友，是伯尔尼发明专利局的局长。格罗斯曼就跟他父亲讲：“你那个朋友不是老想找聪明人到他那里工作吗？你看我那个同学爱因斯坦不就很聪明吗？”从现在的资料来看，在爱因斯坦的老师 and 同学当中，格罗斯曼是第一个看出他聪明的人。结果他父亲真的和那位局长讲了，局长大人就说来面谈吧。一谈，觉得这个年轻人还可以，于是局长说：来吧，给你安排一个工作，三等职员。就是最下等的职员，但是最下等的职员就有一份公务员的薪金，你看现在咱们国家不是也有很多人争着当公务员吗？公务员有“铁饭

碗”啊！于是爱因斯坦就跟米列娃结婚了，建立起一个比较稳定的家庭，他们很快有了两个儿子。

爱因斯坦的科学研究是在去专利局之前开始的，到了专利局以后他的研究继续开展，并开始发表论文。1901年发表一篇论文，1902年两篇论文，1903年一篇论文，1904年一篇论文。论文数量很少，要按照咱们国家现在的规定，这样的人可能都被淘汰了，就这么几篇论文，而且这些论文没有什么特别重要的，都是些毛细管之类的东西。但是这些研究大概对爱因斯坦是个很大的锻炼。1905年是爱因斯坦的丰收年。

为什么要用丰收年这个词？这是因为物理学史上，对牛顿就有一个“丰收年”的说法。牛顿在剑桥大学毕业留校后不久，英国闹鼠疫，于是他躲回家里去了。他23岁到25岁之间有一年半的时间，在他母亲的庄园里度过。按照牛顿后来的说法，他的力学三定律、万有引力定律，以及微积分的构思、对光学的想法，全都是那时候产生的。所以那一年半时间，被称作牛顿的丰收年。

爱因斯坦的丰收年

爱因斯坦于1905年陆续完成了5篇论文。除去一篇博士论文之外，其余4篇都是发表了：3月份提交，6月份发表了光量子说，就是解释光电效应的论文；4月份他把博士论文提交了；然后7月份发表了用分子运动论解释布朗运动的论文；9月份发表了狭义相对论，这篇论文不叫狭义相对论，相对论的名字不是爱因斯坦取的，这篇文章叫《论运动物体的电动力学》；9月份提交、11月份发表了有关的论文。此外，还有一篇是1905年提交，第二年发表的。

$$E = mc^2 \quad (1.2)$$

现在我们来看这5篇论文，除去那篇博士论文以外，其他4篇都是可以得诺贝尔奖的，都是非常硬的文章。我们现在看到很多人得诺贝尔奖，其实他们的贡献究竟是什么一般人也不清楚。即使你稍微知道一

点，过两年也忘了，没有什么太大的发现。而爱因斯坦这几篇论文都非常重要，都影响深远，不是一般获诺贝尔奖的论文比得上的。

专利局——科学发现的摇篮

爱因斯坦的这些杰出工作基本都是在发明专利局做出来的。当他做出成就以后，有的人就开始说：你看，我们的社会有多么不公，爱因斯坦这么伟大的人居然没有一个学校愿意要他，让他在专利局浪费时间！

在这种议论产生以后，他的一位朋友，数学大师希尔伯特说了一句很重要的话：“没有比专利局对爱因斯坦更适合的工作单位了！”为什么呢？就是这个单位事少，清闲！当时德国的学校里老师都得教学，而且教学任务很多、很重，比今天中国重点大学老师的教学任务多多了，讲许多课。另外呢，还要安排科研任务和科研时间，但只能做上面规定的科研题目，不是自己想研究什么就能研究什么。

而爱因斯坦到专利局后，虽然有时要审查一些永动机之类的“发明”，会浪费掉一些时间，可也还有不少空闲时间。于是他把要看的東西摊放在抽屉里，一看领导不在就拽出来钻研，看到领导来了就把抽屉关上。有几次局长注意到爱因斯坦在看本职之外的东西，但局长觉得这个年轻人很爱思考，就不怎么管他。这种宽容的态度和空闲的环境给爱因斯坦创造了科研的条件。当那位开明的局长听说爱因斯坦发表布朗运动这篇论文，证明了分子的存在之后，还马上给他涨了工资。

爱因斯坦大学毕业时确实曾经向很多大学求职，但人家都不要他。当时爱因斯坦怀疑是韦伯捣的鬼，因为那时教授很少，一个大学就一个物理教授，瑞士也没有几个大学，各校的教授都互相认识。他到一所学校去求职，那里的教授肯定会写信问韦伯，说：你的这个学生怎么样啊？爱因斯坦估计韦伯没讲他好话，但此事没有任何证据。前些年有一所大学在整理档案的时候，翻出了当年爱因斯坦的求职信，曾对记者说：“你们看啊，当年爱因斯坦到我们这里求过职，我们没有要他。”

高度评价阿劳中学

爱因斯坦做出成就以后，他曾经回顾大学和中学受教育的境遇。他高度评价了阿劳中学：“这个中学用它的自由精神和那些不依仗外界权势的教师的淳朴热情，培养了我的独立精神和创造精神。正是阿劳中学，成为孕育相对论的土壤。”你看这评价多高，没说他的大学是孕育相对论的土壤，而说这个中学是孕育相对论的土壤。

4. 狭义相对论

好，现在我们来看法义相对论，我简单介绍一下狭义相对论的几个重要成就。狭义相对论建立的基础有两个：一个是相对性原理，就是物理规律在所有的惯性系当中都一样；另外一个足光速不变原理，光速在任何一个惯性系中都是同一个常数 c ，与观测者相对于光源的运动速度无关。

同时的相对性

在这两条原理的基础上爱因斯坦建立起整个理论的框架。从这个框架能得出什么结论呢？一个是“同时”这个概念是相对的。我们都知道，两件事情是不是发生在同一个地点，这个概念是相对的。比如说有一辆电车开过去，电车上有人递给售票员钱，售票员撕了张票给他，这两个动作是否发生在同一地点？车上的人认为“是”，因为两人都没动窝，你给我钱我给你票。但车下的人认为“不是”，这个乘客给钱的时候车还没开，撕票的时候开出去十几米了，两件事不是在同一地点。所以“同时”，即两件事情是不是发生在同一个地点是相对的，这个概念大家都能接受。但是假如说有两个捣乱的小伙子在车上面放炮，一个在车厢前面，一个在车厢后面，一同“咚~”一声炮响，最后把警察找来了解情况，车上的人会说他们两人“同时”点的炮，车下的人会怎么认为呢？当然也会认为是“同时”点的。对不对？但是爱因斯坦的相对论却告诉我们：当电车的速度接近光速的时候，车上的人认为车头车尾“同时”发生的两件事，车下的人就会认为不是在同一时间发生的，这就是“同

时”的相对性。

动钟变慢

另外一个运动的钟会变慢。一个钟如果往前运动，如图1-3所示，比如说我所在的这个参考系，有一列钟，我把它们互相都对准。你所在的参考系，也有一列钟互相对准。这两列钟平行放置，相向运动。这两列钟相对运动的时候，如果我的任何一个指定的钟，跟你的每个钟都只对一次，然后就跑过去了，你那列钟中的任何一个，也与我这列钟的每一个只遭遇一次。那么你会觉得我的指定钟慢了。我也会觉得你的指定钟慢了。如式（1.3）所示，当动钟走过 dt' 时间，静钟走过的时间是 dt 。这是相对论的一个结论：即动钟变慢。

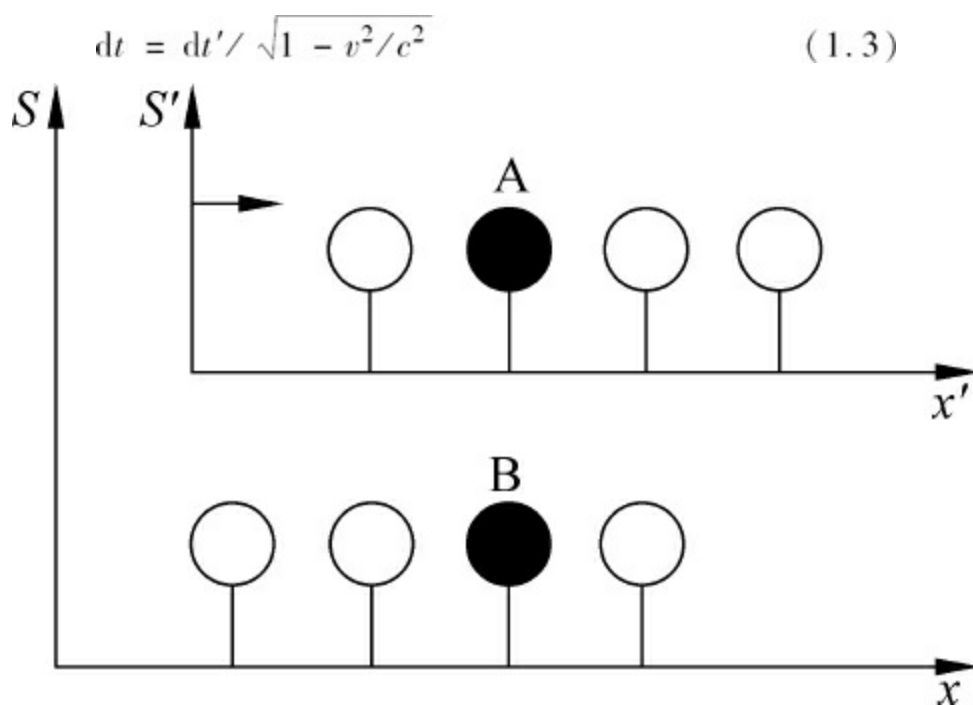


图1-3 动钟变慢

动尺缩短——洛伦兹收缩

同样的，如果双方各有一个尺子，平行放置，相对运动，如图1-4所示。两个尺子这么一下过去，我“同时”量你的尺子就会觉得你的尺子缩短了，你“同时”量我的尺子也会认为我的缩短了。双方都认为对方的

钟慢，对方的尺子缩短。如式（1.4）所示，尺子静止时长度为 l_0 ，以速度 v 运动时，长度缩短为 l 。这也是相对论的一个结论：即动尺缩短，又称洛伦兹收缩（详见本讲附录）。

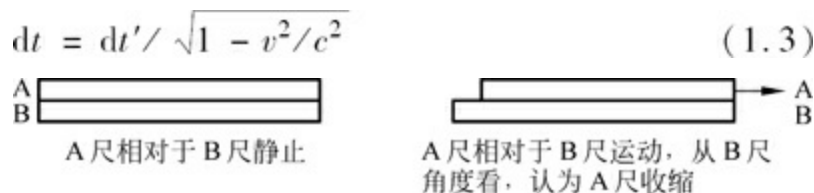


图1-4 动尺缩短

速度叠加

另外就是相对论是禁止超光速的，相对论的速度叠加公式不是我们通常用的、简单的平行四边形法则。比如说有一列火车（见图1-5），它的速度是 v ，有一个人在火车顶上以速度 u' 跑，那么总的速度是多少呢？相对于地面的速度是多少呢？有人以为就是 $u'+v$ ，但是相对论的公式是这样一個公式：

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}} \quad (1.5)$$

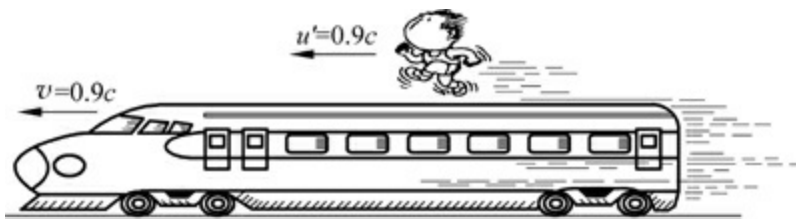


图1-5 速度叠加

这个公式就保证了人和火车跑得再快，即使火车速度达到 $0.9c$ ，上面相对于火车跑的人的速度也达到 $0.9c$ ，但是加在一起不是 $1.8c$ ，而是 $0.9945c$ ，还是小于 c ，再快也超不过光速 c 。

动质量算质量吗？

相对论还有一个公式，就是爱因斯坦那个年代，有人提出了一个动

质量的概念，即式（1.6）所示的动质量 m ，有

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (1.6)$$

就是说一个物体静止的时候质量是 m_0 ，如果它以速度 v 运动的时候，它的质量会增加为 m 。不过，动质量这个概念现在有争议。爱因斯坦等人主张使用动质量和静质量的概念。但是朗道等人认为动质量的概念是不必要的。应该只用静质量，只承认静止的那个质量是真正的质量。

朗道是非常杰出的物理学家，杨振宁先生认为朗道是20世纪三位最伟大的物理学家之一，另两位是爱因斯坦和狄拉克。朗道能对物理学的所有领域发表重要评论。现在有相当大一批物理学家同意取消“动质量”这个概念，但这种观点将会导致只有能量守恒，不存在质量守恒。为什么呢？比如说电子和正电子相撞湮灭了，变成没有静质量只有动质量的光子，但动质量又不算质量，静质量又没有了，这时候质量就不守恒了。所以要牺牲质量守恒这个概念，只有能量守恒。有人说，爱因斯坦本人也同意了“只有静质量才是质量”这个观点，但是爱因斯坦只是在给别人的私人信件中，很婉转地说这个观点是有道理的。爱因斯坦从来没有公开写过文章说只有静质量才算质量，动质量概念应该取消。所以关于这个问题大家会看到有一些争议，有争议也没有有什么关系，说明科学正在发展。大家知道动质量的概念用起来还是比较方便的，很多书现在还在用。

质能关系——质量就是能量

还有就是 $E=mc^2$ ，这个公式是研制原子弹的理论基础之一，它的意思是说任何一个物体都有两种性质，一个是能量，一个是质量。比如说我这里有个茶杯，我说它有能量，但不是指杯中水的热能，水的热能很少，而是指水和茶杯总质量对应的固有能。这个固有能如果一旦全部释放出来，全部转化为热运动能和光能，可以把北京城全炸掉。所以上

面那个公式是研究核能的一个基础。我以后会有单独的讲座专门讲这个问题。

动能表达

还有关于动能的概念，按照相对论，动能应该是动质量对应的能量减去静质量对应的能量。大家看这个公式

$$T = mc^2 - m_0c^2 = m_0c^2 \left(\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right) = \frac{1}{2}m_0v^2 + \frac{3}{8}m_0 \frac{v^4}{c^2} + \dots$$

(1.7)

可是牛顿力学只承认展开的第一项。但当运动速度很高的时候，后边这些高阶项不能忽视，还应该加进来。

5. 神奇的相对论效应

双生子佯谬

好，现在我准备最后讲一下双生子佯谬。这是大家都感兴趣的问题。前面谈到两个人在惯性系中作相对运动。双方都说对方的钟慢了，我说你的钟慢了，你说我的钟慢了。这俩钟是再也不碰面了。有人说让其中一个钟“回来”，可一回来它就要偏离惯性运动，不是惯性系中的钟了。

最初相对论只在惯性系当中讨论问题。但是，法国的物理学家郎之万讨论了一个问题，就是双胞胎兄弟的问题。比如说哥哥坐火箭作星际旅行，绕了一圈以后返回来。返回来后，哥哥好像觉得没过几年，而弟弟已经从年轻人变成一位老头了。真是“天上方七日，地下已千年”了，也就是说，去星际航行的人感觉自己的时间似乎变慢了。这种事情是真的吗？这叫双生子佯谬，谬是错误，佯是假的。佯谬就是假错误，假错误当然就是对的，为什么是这样子呢？后来，曾经有很多人进行过讨论。大家都知道，在相对论当中有个四维时空的概念。就是说除去我们

三维空间以外，还加上时间那一维，就是四维时空。我们每一个人在三维空间中前后左右上下一固定，每个人都是一个点。但是在四维时空当中，由于时间在走，你就会描出一根线来。比如说有一个人他不动，指的是他的空间位置没动，但是他必须跟时间一起走，他要随时间发展往前走。有人说我不走，坚持为一个点，那不行。这是不以人的意志为转移的，必须“与时俱进”。如果你在运动，那么你空间坐标也就变了。

比如说地球上的这个人，相对于星际航行的话，地球就算不动了，那么他描出来的线就是这条A线，如图1-6所示。星际航行的那个人呢，他先离开了地球，然后又返回来，就是这条B曲线。相对论把这种四维时空中的曲线叫做世界线，每一个观测者经历的时间就是他世界线的长度。你看，留在地球上的人的世界线是A，出去的人的世界线是B，两条世界线的长度显然不一样。哪个人的世界线长他就老，哪个人的世界线短他就年轻。大家一看，哟，A线比B线短，似乎地球上这个人年轻。你不是说地球上这个人老吗？那是怎么回事啊？你这是上了伪欧几何的当。欧几里得空间我们都知道，斜边的平方等于两条直角边的平方和，可是闵可夫斯基空间是伪欧几里得空间。时间与空间坐标的长度中间差一个负号，不都是正号。因此斜边的平方等于两条直角边的平方差。所以这条B曲线反而比A短，所以星际航行的那个人年轻，地球上这个人岁数比较大。有人问能年轻多少？

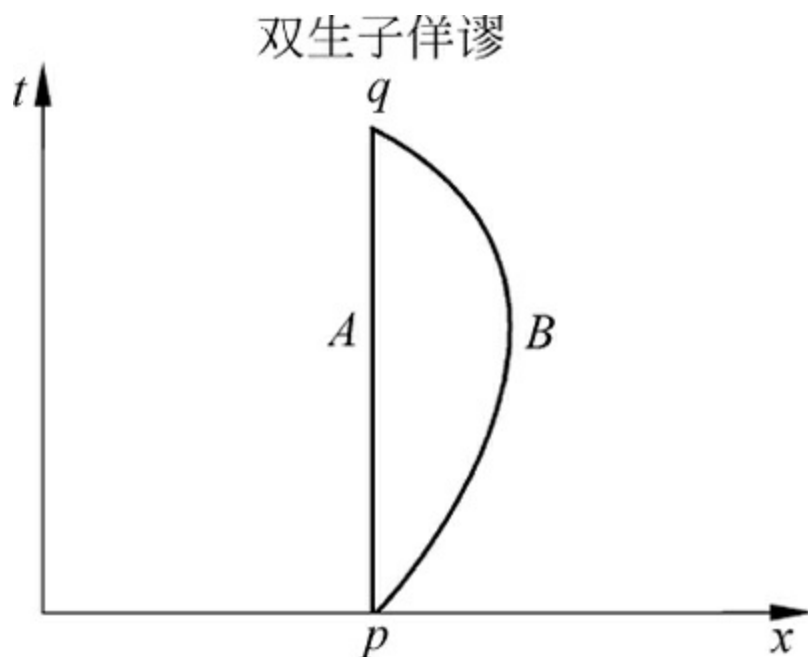


图1-6 双生子佯谬

我给大家举个例子，比如说有人去比邻星旅行，比邻星是除去太阳以外离我们最近的一颗恒星，有多远呢？四光年，就是说光走四年就到了，很近。如果有人坐火箭去这颗星旅行，如果他是三倍的重力加速度加速；有人说我以无穷倍的重力加速度加速行不行？无穷倍不行，一下子就把你压扁了。星际航行的宇航员加速时承受的重力很大，你看杨利伟当时起飞的时候，不是有一段时间他都觉得身体要坚持不住了吗？就是因为重力加速度非常大，一般人承受不了。现在研究认为，三倍的重力加速度还可以勉强凑合。所以就以三倍的重力加速度加速，加速到每秒25万公里以后，就改为惯性运动，关闭发动机出现失重现象。待接近比邻星后，再以三倍的重力加速度减速，直至在比邻星附近的行星上降落。这时必须减速，你不减速就撞上去了，是吧？返回时以同样的方式返回。这样的话，如果有个宇航员坐火箭去了比邻星的行星一趟，火箭上的人觉得往返一共用了7年，而地球上的人觉得他走了多长时间呢？走了12年。地球上的兄弟A可以感觉到自己已经比同胞兄弟B老了。

不过，这还不算老得很明显。假如有人想到银河系中心去旅行，我们的地球不在银河系中心，位于偏离银河系中心约2.8万光年的宇宙中。银河系的直径有10万光年的样子，半径是5万光年。从地球到银河系中心附近，距离大概有3万光年。设想有人坐火箭到银河系中心附近的一颗行星去旅行，然后再返回来。设计的方案是这样：由于时间太长了，就用两倍的重力加速度而且一直维持不变。如果用三倍的重力加速度然后再失重，火箭中的人可能更受不了。假如长期是两倍的重力加速度可能还好受一点。那么就以两倍的重力加速度加速，加速到距目的地中点的时候，再以两倍的重力加速度减速到达那颗星。然后采用同样的方式回来，这时飞船上的人经过了多少年呢？飞船上的人一共经过40年，这还可以，是吧？二十岁的小伙子走了，回来六十岁，还行。那么地球上已过了多少年呢？地球上已过了6万年！所以如果有人完成这样一次旅行的话，地球上的人肯定要开一个盛大的庆祝会，欢迎自己6万年前的祖宗回来了。我讲的这些是有科学依据的，都是用相对论严格计算出来的。

星际飞船上看到的奇景

另外我还想谈一个问题，除去双生子佯谬之外，星际飞船上的宇航员还会看到什么景象，感受到哪些相对论效应呢？

高速飞行的星际飞船上的宇航员还会看到两种景象，一种是多普勒效应造成的，另一种是光行差效应造成的。

由于多普勒效应，飞船前方的星体射来的光会发生蓝移，后方和侧面星体射来的光会发生红移。因此，宇航员觉得前方的星体颜色变蓝，后方的星体颜色变红。侧面的星体由于横向多普勒效应，也会略微变红。

光行差效应会使宇航员觉得侧面的星体向正前方聚集，后面的星体移向自己的侧面。总之，正前方好像是一个“吸引”中心，随着飞船速度的增加，所有的星体都向那里集中，后方的星体越来越少。从地球起

飞，正在远离太阳系的飞船上的宇航员，会觉得太阳系不在飞船的正后方，而在侧后方，飞船越接近光速，太阳系看起来越远离正后方，随着飞船速度的增加，太阳系从自己的侧面向侧前方移动。当飞船的速度非常接近光速时，他将看到太阳系处于自己的侧前方，飞船的后方已经没有任何星体了。飞船正在逃离太阳系，而在宇航员看来，太阳系不是位于飞船的后方，而是位于侧前方，这是多么奇妙的情景啊！

图1-7显示当宇宙飞船向北极星飞去时宇航员看到的景象。当飞船速度远小于光速时，宇航员看到的天象与地面上的人看到的相同，北极星位于正前方，北斗、仙后等星座围绕着它，南天的星座都看不到。当速度达到光速的一半时，飞行员前方的景象大大变化了，北极星周围的星座都在向中央趋近，挤到虚线范围以内，原来出现在飞船后面的天蝎座和天狼星（大犬座 α 星）也都进入前方的视野。当飞船速度加快到 $0.9c$ 时，南天的十字座和老人星等（这些位于南天的星，生活在地球北半球的人原本看不到）也出现在前方了。飞船速度再进一步趋近光速时，整个南天的星系就都挤到前面去了。

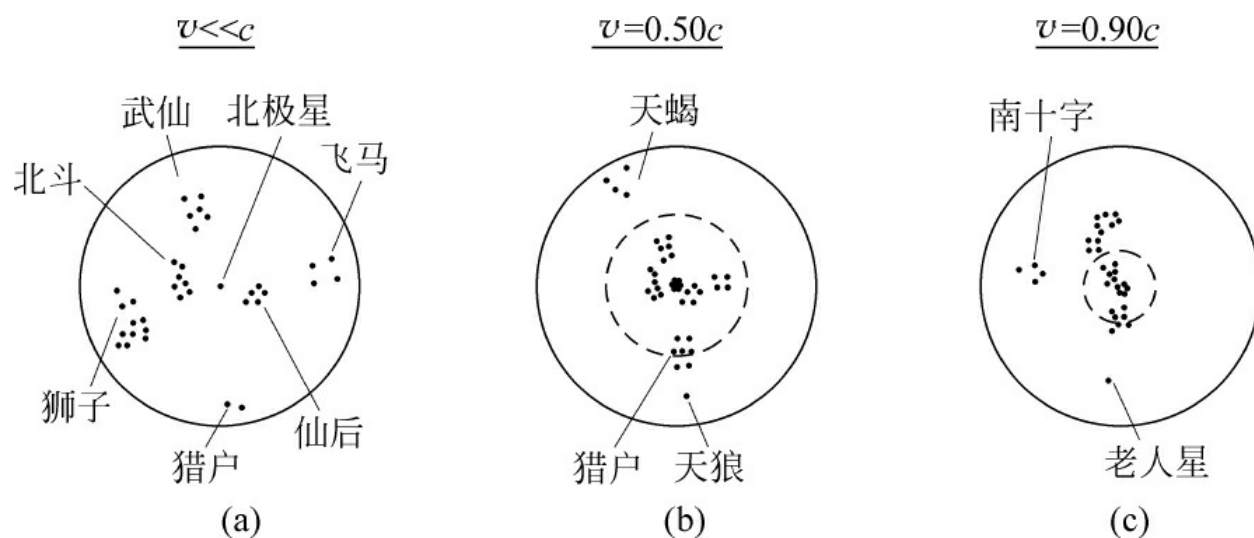


图1-7 飞船宇航员看到的景象

在本讲附录的图1-9和图1-10中，我们曾用打雨伞的人和接雨水的桶

来比喻天文学中的光行差现象。从中容易理解，在运动观测者看来，光线（即图中的雨滴）的来源方向会向自己的正前方聚集。所以，高速飞行的飞船上的宇航员，会观察到所有星系都向正前方汇聚的现象。

上述多普勒效应和光行差现象与飞船发动机是否关闭，飞船是否作加速运动无关，只与飞船的运动速度有关。

宇航员除去看到上述两种景象之外，还会感受到其他一些相对论效应，例如失重和双生子佯谬造成的效应。

当飞船关闭发动机、加速度为零时，宇航员会处于完全失重的状态，这时飞船作惯性运动飞行（见第二讲）。当飞船加速时，宇航员将感受到惯性力，飞船转动时，他们将感受到惯性离心力和科里奥利力。由于等效原理，在飞船那样狭小的空间区域内，飞行员无法区分这些惯性效应造成的力和万有引力，因此加速度和转动形成的惯性力，可以视作人造重力来加以利用。例如，在未来的星际航行中，可以制造人造重力来缓解长期失重给宇航员生理机能带来的不利影响。

今天我就讲这么多。我想问问大家有没有什么问题，有没有？勇敢点提出来，没关系。（现场无人提问题）。

20世纪20年代，德国的哥丁根大学有一个优秀青年组成的“物质结构研讨班”，是由玻恩领导，希尔伯特参与的。研讨班对量子力学的建立和发展做出过重大贡献，并培养了大批第一流人才，例如泡利、海森堡、奥本海默、狄拉克、康普顿等人。这个班有句名言：“愚蠢的问题不仅允许，而且是受欢迎的。”不要怕闹笑话。（现场还是无人提问题）。

杨振宁先生初到美国的时候，有一次，一位美国的物理学家做报告，讲完了以后青年杨振宁没有太听懂，他就问了一个问题，那个美国教授回答了他，随后别人都没有提问题。“哎呀，”杨振宁想，“别人都听懂了，就我没有听懂啊！真是丢面子。”待了一会儿就听见讲课的教授跟主持人讲：“今天的报告特别失败，除去那个中国人听懂了一部分

以外，别的人都没听懂。”呵呵，所以大家应该勇敢地问。有没有问题？

问：老师，我想问一个问题。刚才那个双生子佯谬，为什么这个问题我们要用伪欧时空处理呢？斜边的平方等于两条直角边的平方和或差，什么时候用“和”，什么时候用“差”？

答：凡是四维时空都要用伪欧的，因为时间那一项的正负号是跟空间相反的。如果你没有用时间，全是空间坐标的就都是加号，斜边的平方就等于两条直角边的平方和。一旦是四维时空，把时间加进来了，就一定有一个减号，斜边的平方就等于两条直角边的平方差。

问：你说火箭上的人出去转了一圈，他会年轻，那在火箭上的人看来自个儿没动，地球在外头转了一圈回来了。对不对？是不是应该是留在地球上的人年轻，火箭上的人老啊？这个事情是不是应该是相对的？

答：不是相对的而是绝对的。因为火箭上的人真实地感受到了加速，感受到了惯性力。感不感受到惯性力，是真加速和假加速的一个分界线。火箭上的人真加速了，地球上的人没有，所以这是绝对的结果。

第一讲附录 狭义相对论的创立

1. 相对论诞生前夜“以太理论”带来的实验困难

1801年，托马斯·杨的双缝干涉实验表明，光是一种波动。大家都知道，水波的载体是水，声波的载体是空气或其他气态、液态、固态的物质。光既然是波，应该有一种载体。人们想起了古希腊哲学家亚里士多德的以太理论。

亚里士多德主张地球是宇宙的中心。月亮、太阳、水星、金星等天体都围绕地球转动，天体中离地球最近的是月亮。他认为“月下世界”由土、水、火、气4种元素组成，它们组成的万物都是会腐朽的。而比月亮离地球更远的“月上世界”是永恒不变的，充满了轻而透明的“以太”。不过亚里士多德认为，以太只存在于“月上世界”。19世纪的学者们则进一步认为：以太充斥全宇宙。他们认为光就是以太的弹性振动，也就是说光波的载体就是以太。光能从遥远的星体传播到地球，表明以太不仅透明而且弹性极好。

相对论诞生前夜，实验观测引发了与以太理论有关的矛盾。

既然光波是以太的弹性振动，那么以太相对于地球是否运动？当时哥白尼的“日心说”已经被普遍接受，地球不是宇宙的中心。如果认为以太整体相对于地球静止，就等于倒退回“地心说”，大家无法接受这种看法。科学界认为比较合理的设想是：以太相对于牛顿所说的“绝对空间”静止，因而在绝对空间中运动的地球，应该在以太中穿行。这就是说，以太相对于地球应该有一个“漂移”速度。

天文学上的“光行差”现象似乎支持存在以太漂移。然而，迈克耳孙的精确实验却没有测到以太相对于地球的“漂移”速度。也就是说，作为介质的地球似乎带动了周围的以太跟自己一起运动。光行差现象认为地球（介质）运动没有带动以太，迈克耳孙实验又认为带动了以太，这一

观测上的重大矛盾，就是开尔文勋爵在1900年英国皇家学会迎接新世纪的庆祝会上所谈的，物理学的两朵乌云中的一朵。

此外，斐佐的流水实验表明“流水”（运动介质）似乎部分地带动了以太，但又没有完全带动。

总之，光行差现象表明运动介质没有带动以太，迈克耳孙实验表明运动介质完全带动了以太（即以太相对于介质静止），斐佐实验则表明运动介质部分地带动了以太，而又没有完全带动。这三个实验的结论相互矛盾。

洛伦兹等众多物理学家注意的是迈克耳孙实验与光行差现象的矛盾。爱因斯坦注意的则是斐佐实验与光行差现象的矛盾。应该说，这两个矛盾都可以引导人们去创建相对论。

光行差现象

所谓“光行差”效应（即光行差现象），是天文学家早就注意到的一种现象：观测同一恒星的望远镜的倾角，要随季节作规律性变化（图1-8）。

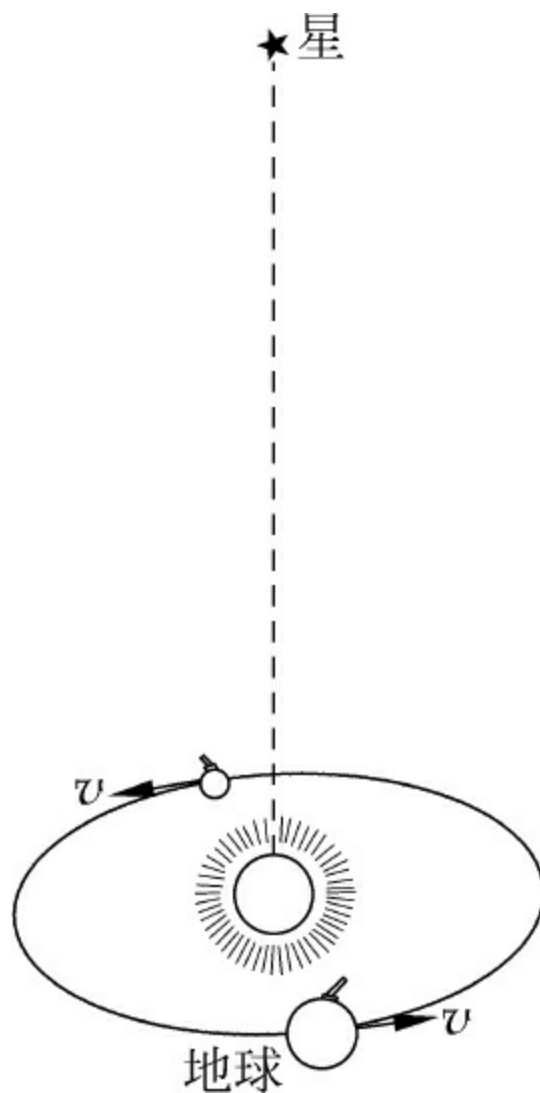


图1-8 光行差现象

此现象很容易理解。比如，不刮风的下雨天，空气不流动，雨滴在空气中垂直下落，站立不动的人应该竖直打伞，跑动的人则应该把伞向跑动的方向倾斜，因为奔跑时空气相对于人运动，形成迎面而来的风，所以雨滴相对于他不再竖直下落，而是斜飘下来。（图1-9）如果有人想接雨水，无风时他应该把桶静止竖直放置。如果他抱着桶跑，则必须让桶向运动方向倾斜，雨滴才会落入桶中（图1-10）。

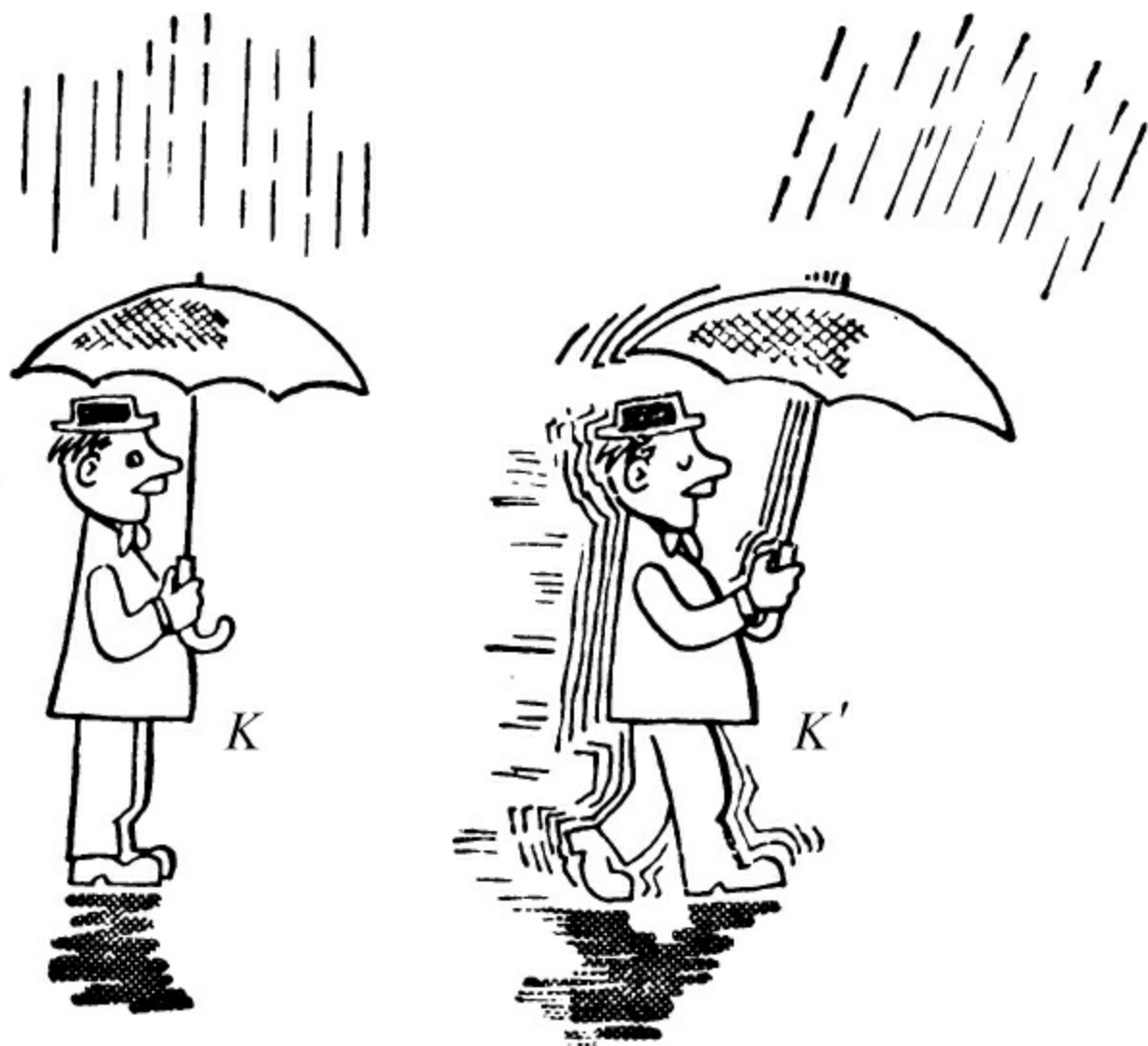


图1-9 雨中打伞

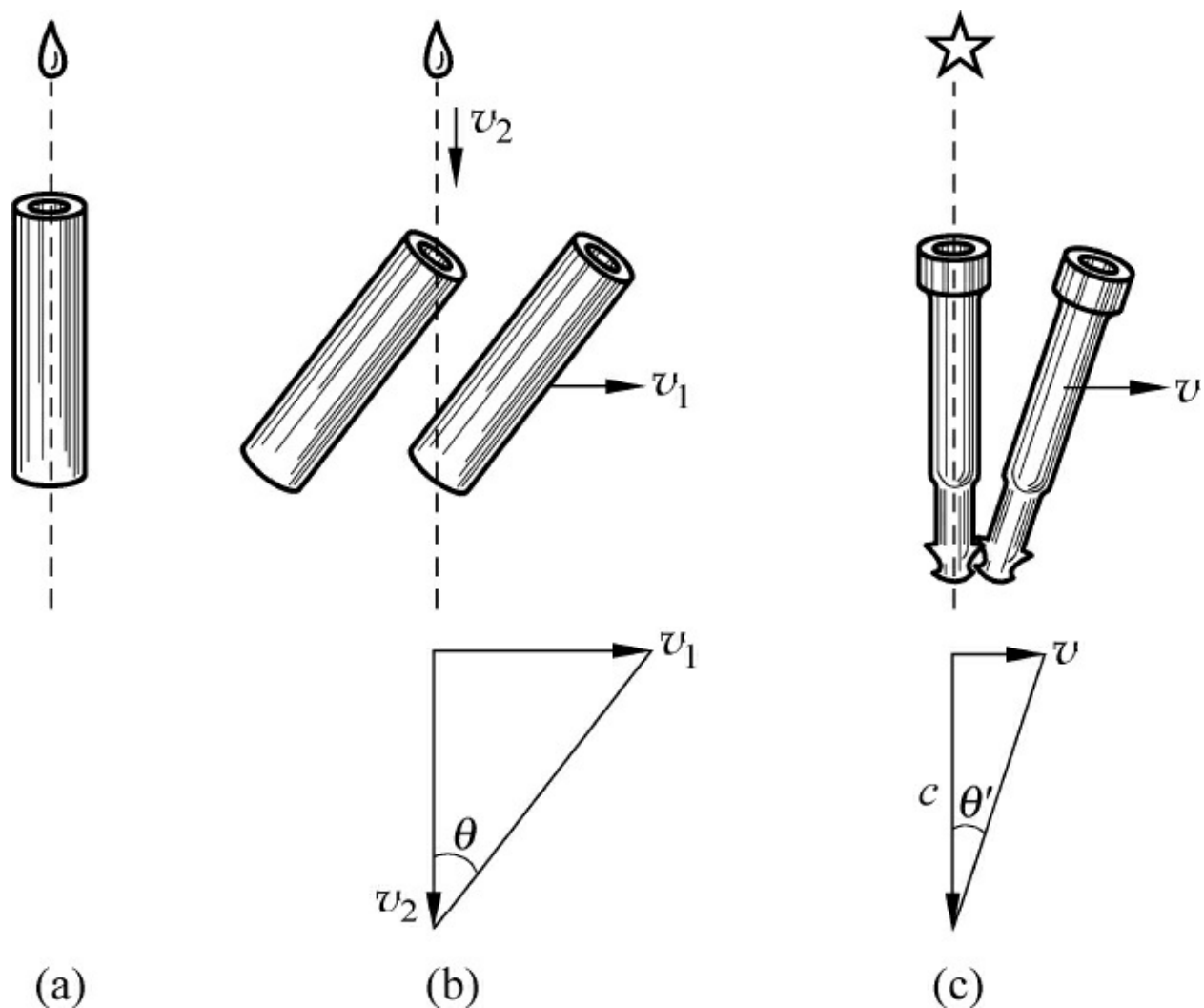


图1-10 接雨水的桶

恒星距离我们十分遥远（除太阳外，最近的恒星离我们也在4光年以上），从它们射来的光可以近似看作平行光。星光在以太中运动，就像空气中的雨滴一样。如果地球相对于以太整体静止，望远镜只须一直指向星体的方向看就可以了。然而地球在绕日公转，地球上的望远镜就像运动者手中的雨伞和水桶一样，必须随着地球运动方向的改变而改变倾角，才能保证所观测恒星的光总是落入望远镜筒内。

“光行差”现象早在1728年就已发现，1810年又被进一步确认，此现象似乎表明地球在以太中穿行。当时科学界认为以太相对于“绝对空间”静止，因此地球相对于以太的速度也就是相对于“绝对空间”的速

度。人们非常希望精确地知道这一速度，然而“光行差”效应的测量精度不够高，于是美国科学家迈克耳孙试图用干涉仪来精确测量地球相对于以太的运动速度。

迈克耳孙实验引来的乌云

迈克耳孙干涉实验如图1-11所示，A为光源，D为半透明半反射的玻璃片。入射到D上的光线分成两束，一束穿过D片到达反射镜 M_1 ，然后反射回D，再被D反射到达观测镜筒T。另一束被D反射到反射镜 M_2 ，再从 M_2 反射回来，穿过D片到达观测镜筒T。把此装置水平放置， v 为以太漂移方向（与地球公转方向相反）。DM $_1$ 沿着以太漂移的方向，DM $_2$ 与以太漂移方向垂直。

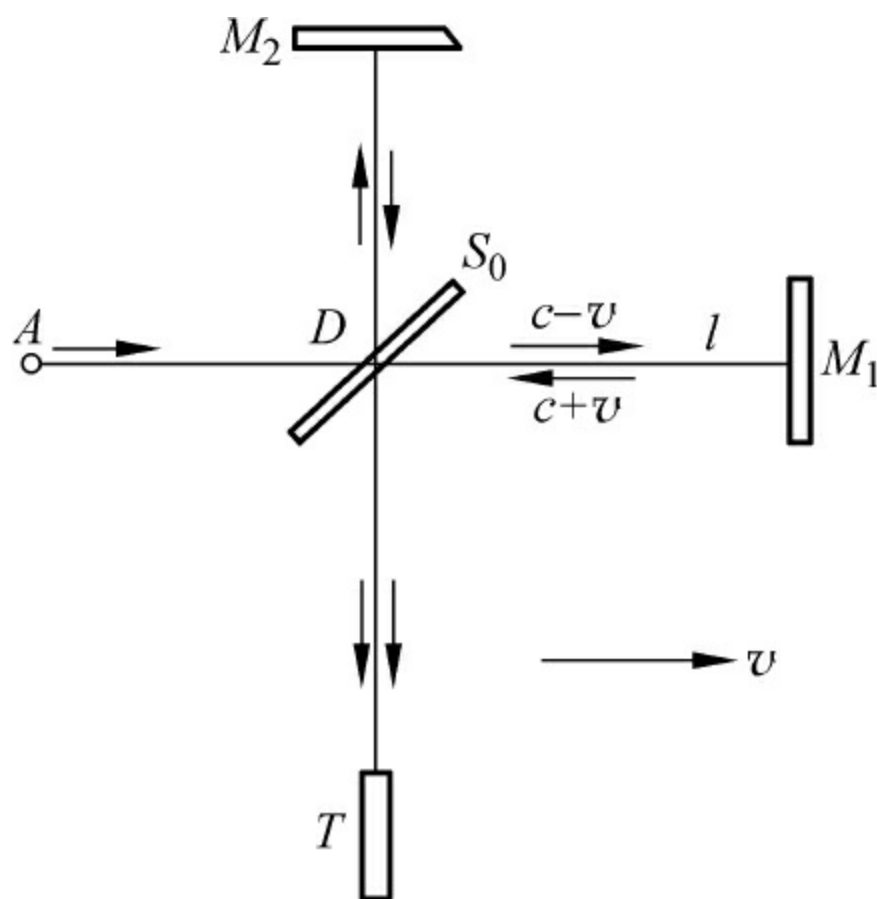


图1-11 迈克耳孙干涉实验示意图

在迈克耳孙干涉装置中运动的光波，就像在河中游泳的人一样。如

图1-12所示，河水以速度 v 相对于河岸流动，河宽 $AB=l_0$ 。一个游泳的人从A出发以速度 u （相对于河水）游到下游 B' 点，再返身以同一速度 u 游回A点， AB' 的长度与河宽相等，即 $AB'=l_0$ 。再让同一游泳者以速度 u （相对于河水）从A出发游向对岸的B点，到达后再以同一速度游回出发点A。但要注意，由于水往下游流，横渡者的游泳方向不能垂直于河岸，那样的话他将被河水往下冲，不可能恰好抵达B点，返回时也会出现同样的情况。为了从A游到B，游泳者游动的方向必须向上游倾斜一个角度，如图1-13所示。所以游泳者垂直渡河的速度应是 $u' = \sqrt{u^2 - v^2}$ 。虽然游泳者横渡的距离与向下游游动的距离都为 l_0 ，但两种情况所需的时间却不同，时间差为

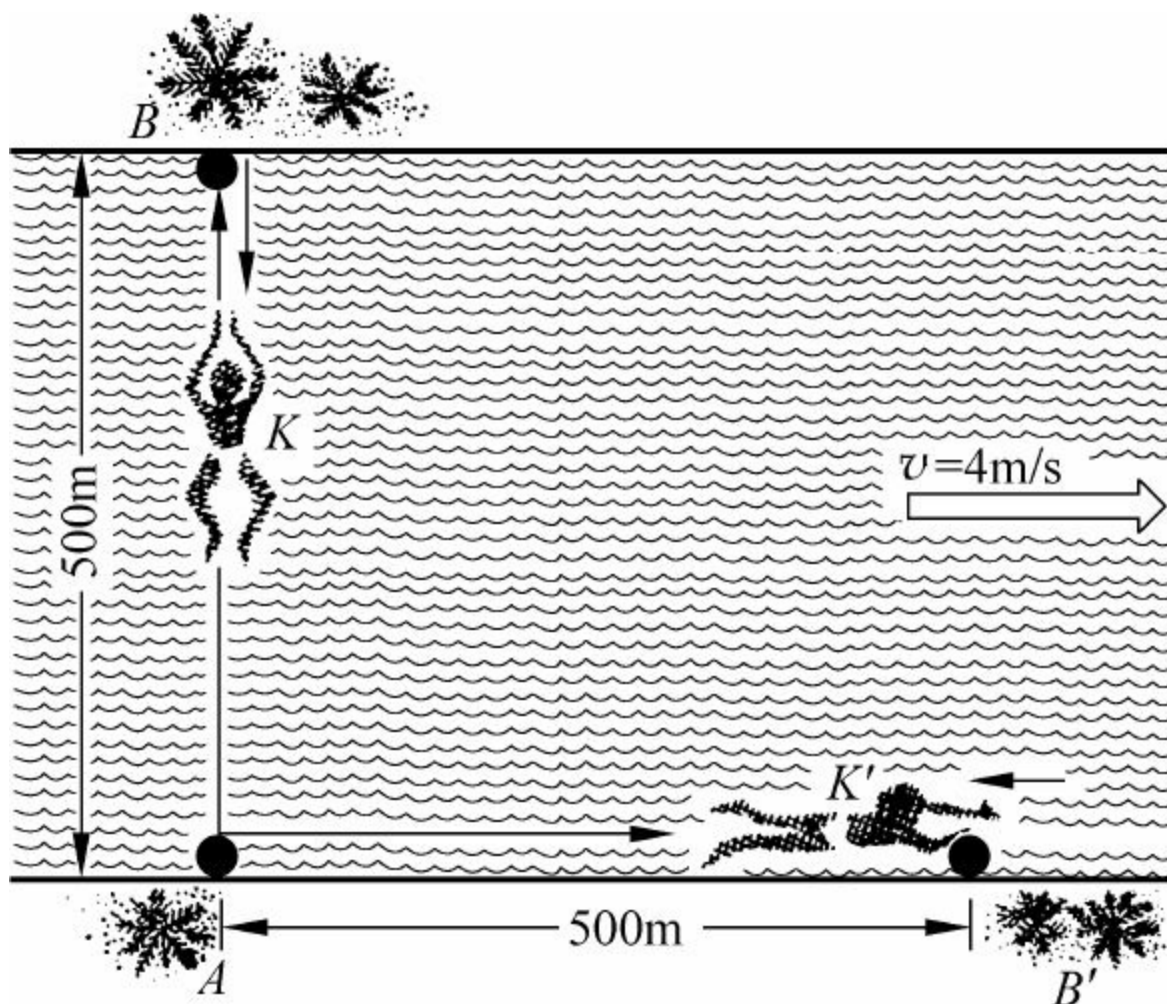


图1-12 在水中游泳的人

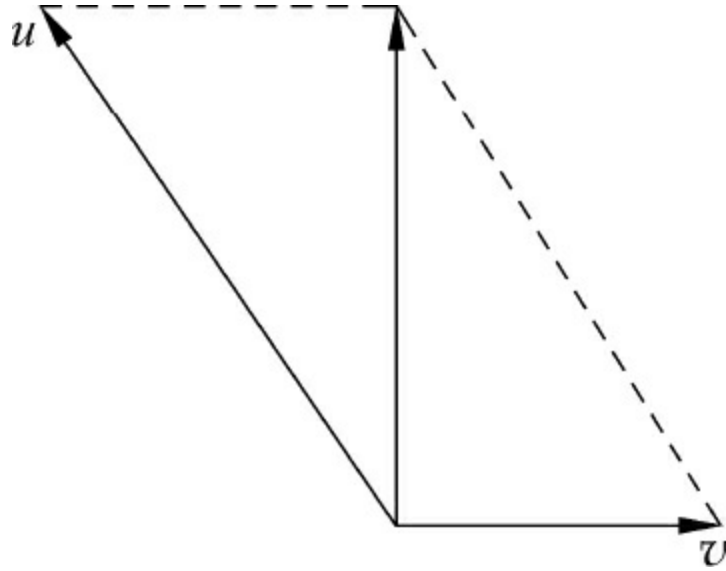


图1-13 渡河速度合成图

$$\Delta t = \frac{l_0}{u+v} + \frac{l_0}{u-v} - \frac{2l_0}{\sqrt{u^2 - v^2}} \quad (1.8)$$

迈克耳孙干涉仪中的光波，就像上面所说的游泳者，河水好比漂移的以太，河岸相当于地球。河水相对于河岸的流动可类比以太相对于地球的漂移。虽然距离DM₁与DM₂相同，但光波经过这两段距离所需的时间却由于以太的漂移而不同，用光波相对以太的速度c取代u，我们用同样的分析可知二者的时间差为

$$\Delta t = \frac{l_0}{c+v} + \frac{l_0}{c-v} - \frac{2l_0}{\sqrt{c^2 - v^2}} \approx \frac{l_0}{c} \left(\frac{v^2}{c^2} \right) \quad (1.9)$$

这就是说；光经过DM₁所需的时间比经过DM₂所需的时间要长。

迈克耳孙把干涉仪在水平面上转90°，让DM₂沿以太漂移的方向，DM₁则垂直以太漂移方向。这时光经过DM₂的时间反而比经过DM₁的时间长。

仪器装置转动90°的结果，将使到达观测镜T的两束光所经历的时间差了2Δt，导致光程差改变

$$2c\Delta t \approx 2l_0 \left(\frac{v^2}{c^2} \right) \quad (1.10)$$

这将引起这两束光形成的干涉条纹产生相应的移动。遗憾的是迈克耳孙没有测出干涉条纹的移动，在误差精度内，条纹的移动是零。迈克耳孙及其助手曾采取多种措施提高实验精度，但结果仍然是零。

“光行差”现象告诉人们以太相对于地球有漂移，迈克耳孙实验则没有测到这种漂移。这就是相对论诞生前夜物理学遇到的一个严重困难，即开尔文所说的乌云中的一朵。

2. 相对论诞生前夜电磁理论引起的理论困难

相对论诞生的前夜，除去以太理论导致的困难之外，物理理论还遇到了另一个困难：麦克斯韦电磁理论似乎与伽利略变换矛盾。

19世纪下半叶，麦克斯韦从介质的弹性理论导出了一组电磁场方程，虽然今天我们知道从介质的振动去推导电磁场方程既不正确也无必要，但麦克斯韦所得的结论还是正确的，他对电磁理论的贡献仍是伟大卓越的。

从麦克斯韦电磁方程组出发，可以得到一个重要结论：电磁波以光速传播，人们很快认识到光波实际上就是电磁波。在电磁理论中，真空中的光速是一个恒定的常数。所谓真空，就是只存在以太，不存在其他介质的空间。伽利略相对性原理告诉我们，力学规律在一切惯性系中都是相同的（注意，伽利略论证的相对性原理，仅对力学规律而言，因此又被后人称为力学相对性原理）。如果把这一相对性原理加以推广，使之对电磁学规律也成立，那么麦克斯韦电磁方程组就应在所有惯性系中都一样，这就是说，光速在任何惯性系中都应相同，都应是同一个常数 c 。按照牛顿的观点，所有相对于绝对空间静止或作匀速直线运动的参考系都是惯性系，惯性系之间可以差一个相对运动速度 \mathbf{v} 。依照速度（矢量）叠加的平行四边形法则，电磁波（即光波）的速度如果在惯性

系A中是 c ，那么，在相对于A以速度 v 运动的另一个惯性系B中，就不应再是 c 了。当 c 与 v 反向时应是 $c+v$ ，而当 c 与 v 同向时，则应是 $c-v$ 。但是，麦克斯韦电磁理论明确无误地告诉我们，光速在所有惯性系中都只能是 c ，不能是 $c+v$ 或 $c-v$ 。那么，毛病出在哪里呢？

回顾一下上面的讨论，不难看出，我们用了以下一些原理：

①麦克斯韦电磁理论，它要求真空中的光速只能是常数 c ；

②相对性原理，它要求包括电磁理论在内的所有物理规律在一切惯性系中都相同；

③伽利略变换，即作为速度迭加原理的平行四边形法则，它被当作伽利略相对性原理的数学体现。

就是这三条原理导致了上述矛盾。

挽救以太理论的尝试：**2**伦兹—斐兹杰惹收缩

相对论诞生之前，“以太”理论在人们的头脑中根深蒂固，虽然物理理论遇到了重大困难，而且迈克耳孙实验与光行差实验也暴露出深刻的矛盾，绝大多数人（包括洛伦兹、庞加莱这样的物理、数学大师）仍然不怀疑以太的存在，不怀疑“光波是以太的弹性振动”。

为了保留以太理论，同时克服上述理论困难和实验困难，当时最杰出的电磁学专家洛伦兹决定放弃相对性原理。他想保留麦克斯韦电磁理论，同时解决迈克耳孙实验与光行差实验的矛盾。为此，他提出，以太相对于绝对空间是静止的。麦克斯韦电磁理论只在相对于以太（即绝对空间）静止的惯性系中成立。光波相对于以太（绝对空间）的速度是 c ，相对于运动系的速度不再是 c 。他又提出一个新效应：相对于绝对空间运动的刚尺，会在运动方向上产生收缩

$$l = l_0 \sqrt{1 - \frac{v^2}{c^2}} \quad (1.11)$$

这一收缩被称为洛伦兹收缩。式中 l_0 是刚尺相对于绝对空间静止时的长

度， l 是刚尺相对于绝对空间以速度 v 运动时的长度， c 是真空中光速。洛伦兹等人认为这种“收缩”是物理学家以前不知道的一种新的物理效应。此效应可以解释为何迈克耳孙实验观测不到地球相对于以太的运动。这是因为沿运动方向放置的干涉仪的臂长发生了洛伦兹收缩，缩短了光程，这一效应抵消了地球相对以太运动带来的光程改变。

$$\Delta t = \frac{l}{c+v} + \frac{l}{c-v} - \frac{2l_0}{\sqrt{c^2 - v^2}} = 0 \quad (1.12)$$

他们认为洛伦兹收缩是物理的，会引起收缩物体内部结构和物理性质的变化。

需要说明的是，洛伦兹是1892年提出上述收缩假设的，爱尔兰物理学家斐兹杰惹声称自己早在1889年就提出了这一收缩假设，并开始在课堂上对学生讲授。然而当时大家看到的斐兹杰惹的有关论文最早是1893年发表的，晚于洛伦兹。斐兹杰惹去世后，他的学生为了给自己的老师讨个公道，翻查各种文献，终于在英国出版的《科学》杂志上查到了1889年斐兹杰惹投给该刊的讨论这一收缩的论文。由于斐兹杰惹投稿给《科学》不久，该刊就倒闭了，斐兹杰惹以为自己的文章没有登出来，事实上此文登在了该刊倒闭前的倒数第二期上。看来，斐兹杰惹发现这一收缩确实早于洛伦兹。所以洛伦兹收缩应该称为洛伦兹—斐兹杰惹收缩。

经典理论的改良：洛伦兹变换的提出

洛伦兹等人进一步认为，作为力学相对性原理数学体现的伽利略变换

$$\begin{cases} x' = x - vt \\ y' = y \\ z' = z \\ t' = t \end{cases} \quad (1.13)$$

应当放弃，而代之以新变换（庞加莱称其为洛伦兹变换）

$$\begin{cases} x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \\ y' = y \\ z' = z \\ t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}} \end{cases} \quad (1.14)$$

式中 (x, y, z, t) 为一个指定的事件在相对于以太（即绝对空间）静止的惯性系中的空间坐标和时间坐标， (x', y', z', t') 为同一个事件在运动惯性系中的空间和时间坐标。 x' 轴与 x 轴重合， y' 轴与 y 轴、 z' 轴与 z 轴分别平行，运动方向沿 x 轴。 v 是运动系相对于静止系（绝对空间）的速度， c 是光速。这里，除去公式上的数学差异外，物理上还有一个重要区别：式（1.13）表示的是任意两个惯性系之间的变换，式（1.14）表示的是惯性系相对于绝对空间的变换。即式（1.13）中的速度 v 只是两个惯性系之间的相对速度，与绝对空间无关。而式（1.14）中的 v 却是惯性系相对于绝对空间的绝对速度。式（1.14）中的 (x, y, z, t) 特指相

对于绝对空间静止的惯性系的空间和时间坐标。

从洛伦兹变换可以推出刚尺收缩公式（1.11）。而且麦克斯韦电磁方程在洛伦兹变换下形式不变（不过，洛伦兹认为，用洛伦兹变换算得的、用运动坐标系标出的电磁量及其他物理量或几何量，都没有测量意义，因而不能看作是真实的量，只是一种表观的量）。伽利略变换不具备这两个优点。洛伦兹等人用式（1.11）和式（1.14）克服了迈克耳孙实验造成的困难，代价是抛弃了相对性原理。

需要补充说明，佛格特早在1887年就提出了类似于洛伦兹变换的变换，但有错误。洛伦兹知道佛格特的工作，但没有足够注意。首先给出洛伦兹变换正确形式的是英国物理学家拉摩，他于1898年给出了这一变换。后来斐兹杰惹也独立给出了洛伦兹变换的正确形式。而洛伦兹本人则是在1904年发表这一变换的，上述事实表明，一个重要的科学结论，在条件接近成熟的时候，往往会被许多学者分别独立地多次发现。

3. 走向狭义相对论

爱因斯坦独辟蹊径

爱因斯坦没有注意洛伦兹等人的工作，也没有注意迈克耳孙实验，他主要抓住的是斐索实验与光行差实验的矛盾。光行差与迈克耳孙实验的矛盾体现在运动介质是否拖动以太上。光行差现象表明，作为介质的地球完全没有拖动以太；迈克耳孙实验则表明，似乎地球完全拖动了附近的以太。斐索实验研究了流水对光速的影响，其结论是作为介质的流水似乎部分地拖动了以太，但又没有完全拖动。这也与光行差现象认为运动介质完全不拖动以太的结论相冲突。爱因斯坦认识到解决上述矛盾最简单的方法就是放弃以太理论，不承认有以太存在。

爱因斯坦深受奥地利物理学家兼哲学家马赫影响。他阅读过马赫的著作《力学史评》，在这本书中，马赫勇敢地批判占统治地位的牛顿的绝对时空观，认为根本就不存在绝对空间和绝对运动，也不存在以太，

一切运动都是相对的。爱因斯坦接受马赫相对运动的思想，认为观测不到的东西都不应该轻易相信其存在，哪个实验证明了存在绝对空间？谁看见过以太？因此以太理论和绝对空间概念都应该放弃。他认为伽利略变换不等于相对性原理。他考虑了①麦克斯韦电磁理论（包括真空中的光速 c 是常数的结论），②相对性原理与③伽利略变换之间的矛盾，认为“麦克斯韦电磁理论”和“相对性原理”比伽利略变换更基本。他认识到，如果既坚持“相对性原理”又坚持“麦克斯韦电磁理论”，就必须承认真空中的光速在所有惯性系中都是同一个常数 c ，即必须承认“光速不变”。他把“光速不变”看作一条基本原理，称为“光速不变原理”。注意，“光速不变原理”不是说在同一惯性系里真空中的光速处处均匀各向同性，是一个常数 c ，而是说在任何惯性系中测量，真空中的光速都是同一个常数 c ，光速与光源相对于观测者的运动速度无关。

爱因斯坦得出光速不变原理不是偶然的，而是经历了长时间的思考过程。

他在阿劳中学学习时就考虑过一个思想实验：假如一个观测者以光速运动，追光，这个观测者应该看到一个不依赖于时间的波场。但是谁都没有见过这种情况。这个有趣的问题表明，人似乎不可能追上光，光相对于观测者似乎不会静止，一定有运动速度，通常的速度叠加法则好像对光的传播问题不适用。这个思想实验不时浮现在爱因斯坦的脑海中。

此外，爱因斯坦知道，天文望远镜对双星轨道的观测（图1-14），支持光速与光源运动无关的观点。如果光速与光源运动速度有关，双星中向着我们运动（趋近）的那颗星和背离我们运动（远离）的那颗星发出的光，飞向地球的速度将不同。这将导致两颗星同时发出的光会一先一后到达我们眼中；或者说我们同时看见的这两颗星的图像，产生的时间不是同一时刻。如果真是这样，我们看到的双星轨道应该产生畸变。但天文观测没有发现这种畸变，双星轨道是正常的椭圆。这支持了光速

与光源运动速度无关的看法。

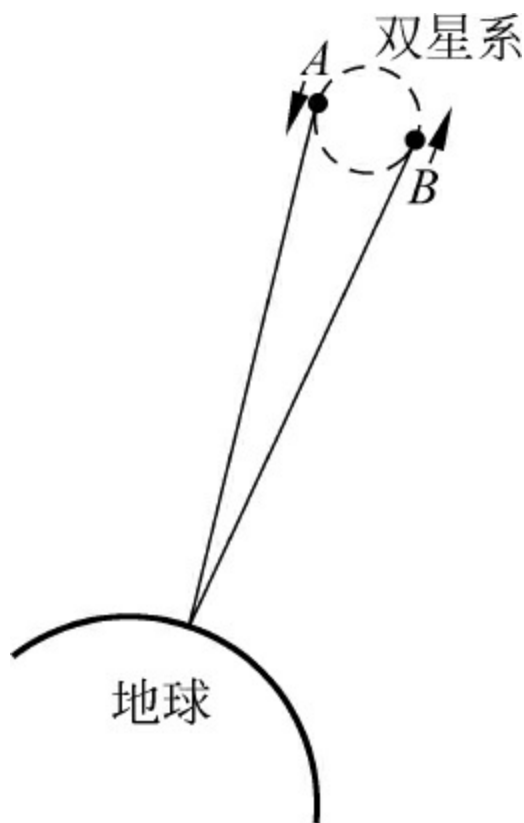


图1-14 对双星轨道的观测

经过长时期的思考后，爱因斯坦终于解开了这个难解之谜。他认识到速度叠加法则并非物理学的根本原理，这个法则也不等价于“相对性原理”的数学表达。“光速的绝对性”（即光在所有惯性系中的速度都是同一个常数 c ）才是一条应该坚持的基本原理，他称其为“光速不变原理”，并把“光速不变原理”和“相对性原理”一起，作为自己的新理论（相对论）的基石。

爱因斯坦是在长时间的反复思考之后，才得出这一原理的。早在他的相对论论文发表之前一年多，他就认识到相对性原理和麦克斯韦电磁理论都是大量实验证实的理论，都应该坚持。但这样导致的“光速不变”结论似乎与建立在伽利略变换基础上的速度叠加法则以及人们的日常观念相矛盾，爱因斯坦觉得“这真是个难解之谜”。

1905年5月的一天，他带着这一问题专门拜访了他的好友贝索（“奥林匹亚”科学院的一个成员）。经过一下午的讨论，爱因斯坦突然明白了，问题出在“时间”上，通常的时间概念值得怀疑。“时间并不是绝对确定的，而是在时间与信号速度之间有着不可分割的联系。有了这个概念，前面的疑难也就迎刃而解了。”他认识到如果坚持把相对性原理和光速不变（即光速与观测者相对于光源的运动速度无关）都看作公理，异地时钟的“同时”将是一个相对的概念。5周之后，爱因斯坦开创相对论的论文就寄给了杂志社。

贝索是个一事无成者的典型。他一生都在听课、学习，课听了一门又一门，书学了一本又一本。他还喜欢与别人争论，反驳别人的意见，但从不想自己去完成一件独立的工作。这次与爱因斯坦的讨论，大大地启发了爱因斯坦，但他自己并未搞清启发了爱因斯坦什么。当爱因斯坦感谢他在讨论中帮助了自己时，他感到茫然。爱因斯坦在这篇创建相对论的划时代论文的最后感谢了贝索对自己的帮助和有价值的建议。贝索十分激动，说“阿尔伯特，你把我带进了历史”。

爱因斯坦1922年在日本京都的一次演讲中曾提到他与贝索的这次讨论。讨论使他认识到两个地点的钟“同时”，并不像人们通常想象的那样，是一个“绝对”的概念。物理学中的概念都必须在实验中可测量，“同时”这个概念也不例外。而要使“同时”的定义是可测量的，就必须对信号传播速度事先要有一个约定。由于真空中的光速在电磁学中处于核心地位，爱因斯坦猜测应该约定（或者说“规定”）真空中的光速各向同性而且是一个常数，在此基础上来校准两个异地的时钟，即定义异地时间的同时。研究表明，在约定光速并承认光速的绝对性（光速不变原理）的基础上定义的“同时”将是一个相对的概念。我们看到，定义两个地点的钟同时，必须首先约定光速各向同性而且是一个常数。要在作相对运动的所有惯性系中，都用对光速的同一个约定来定义异地时钟的“同时”，则必须假定光速是绝对的。爱因斯坦曾经与贝索等人一起阅

读过庞加莱的《科学与假设》，该书就议论过时间测量与光速的内在联系。庞加莱猜测，要测量时间，要校准不同地点的钟，可能首先要对光速有一个约定。与贝索的讨论可能使爱因斯坦想起了庞加莱的观点，不过爱因斯坦未明确指出这一点。此外，与贝索的讨论还可能再次使爱因斯坦想到了他在阿劳中学读书时考虑过的那个思想试验：以光速运动的观测者将看到光是不依赖于时间的波场，但从未有人见过这种情况，所以比较自然的想法是，光不可能相对任何观测者静止，对任何观测者都一定作相对运动。

爱因斯坦能够从纷乱的理论探讨和实验资料中，认识到应该把光速看作绝对的，并毅然提出这一全新的观念，是极其难能可贵的。在光速不变原理和相对性原理的基础上，他推出了两个惯性系之间的坐标变换关系，这个关系就是洛伦兹等人早已得出的变换公式（1.14）。不过，爱因斯坦是在不知道洛伦兹等人的工作的情况下，独立推出这一公式的。更重要的是，爱因斯坦对公式（1.14）的解释与洛伦兹完全不同。洛伦兹认为相对性原理不正确，认为存在绝对空间（以太），变换式（1.14）中的速度 v 是相对于绝对空间的，因而，变换式（1.14）描述的是相对于绝对空间运动的惯性系与绝对空间静止系之间的关系。爱因斯坦则认为，相对性原理成立，不存在绝对空间，不存在以太，式（1.14）描述的是任意两个惯性系之间的变换， v 是这两个惯性系之间的相对速度，根本与绝对空间的概念没有关系，所以他赞同把自己的理论叫作相对论。

我们看到非常有趣的情况，相对论的最主要的公式洛伦兹变换，是洛伦兹最先给出的，但相对论的创始人却不是洛伦兹而是爱因斯坦。应该说明，这里不存在篡夺科研成果的问题。洛伦兹本人也认为，相对论是爱因斯坦提出的。在一次洛伦兹主持的讨论会上，他对听众宣布，“现在，请爱因斯坦先生介绍他的相对论。”之所以如此，是因为洛伦兹一度反对相对论，他还曾与爱因斯坦争论过相对论的正确性。特别

有趣的是，“相对论”这个名字不是爱因斯坦起的，而是洛伦兹起的。在争论中，为了区分自己的理论和爱因斯坦的理论，洛伦兹给爱因斯坦的理论起了个名字——相对论。爱因斯坦觉得这个名字与自己的理论还比较相称，于是接受了这一命名。

建立狭义相对论最困难的思想突破

一般介绍相对论的文章都非常强调爱因斯坦之所以能建立相对论，关键是他坚持了“相对性原理”。在当时的情况下，爱因斯坦正确地认识到“相对性原理”是应该坚持的一条根本性原理，并认识到伽利略变换并不等价于“相对性原理”，然后放弃后者而坚持前者，的确是十分不容易的。洛伦兹和大多数物理学家都没有认识到“相对性原理”是最应该坚持的根本性原理。

但是，应该注意到，关于运动相对性的观念自古以来各国都有。到了17世纪，伽利略已经通过对话的形式正确地给出相对性原理的基本内容。牛顿虽然认为存在绝对空间，同时认为转动是绝对运动，但他还是认为各个惯性系是等价的。应该说，牛顿在他的理论中应用了相对性原理。

到了1900年前后，虽然洛伦兹等人考虑放弃相对性原理，但由于马赫对牛顿绝对时空观的勇敢批判，深受马赫影响的爱因斯坦还是比较容易认识到应该坚持“相对性原理”的。

然而，仅仅认识到坚持“相对性原理”，还不足以建立相对论。庞加莱已经正确地阐述了“相对性原理”，并认识到了真空中的光速可能是一个常数，甚至认识到光速可能是极限速度，但是他仍未能建立相对论。这是因为建立相对论还必须实现观念上的另一个更为重要的突破：认识到光速的绝对性，即“光速不变原理”。

爱因斯坦曾明确指出，狭义相对论与（伽利略和牛顿建立的）经典力学都满足相对性原理，“因此，使狭义相对论脱离经典力学的并非相对性原理这一假设，而是光在真空中速度不变的假设。它与狭义相对性

原理相结合，用众所周知的方法推出了同时的相对性，洛伦兹变换及有关运动物体与运动时钟行为的规律。”

这就是说，承认相对性原理，又承认光速绝对性，必将导致时间观念发生根本变化：“同时”这个概念不再是“绝对”的，而是“相对”的了。同时的相对性与人们的日常观念严重冲突，非常不易被接受。所以认识到“光速的绝对性”，进而认识到“同时的相对性”，是建立相对论过程中最困难也最重要的物理思想突破。

爱因斯坦是相对论的唯一缔造者

1905年前后，许多人都已接近相对论（狭义相对论）的发现，在爱因斯坦的论文发表之前，斐兹杰惹和洛伦兹早已提出洛伦兹收缩，佛格特、拉摩、斐兹杰惹、洛伦兹早已给出洛伦兹变换，拉摩已经给出了运动时钟变慢的公式，洛伦兹已经给出了质量公式（1.6），庞加莱已经正确地阐述了相对性原理，并推测真空中的光速可能是常数，而且可能是极限速度。此外，在一些特殊的情况下，质能关系式也已有人探讨。

但是，提出“光速不变原理”的人是爱因斯坦，而不是其他人。正是“光速不变原理”，而不是“相对性原理”，形成了相对论与经典力学的分水岭。另一方面，只有爱因斯坦抛弃了以太理论，从而彻底抛弃了“绝对空间”，因而最彻底地坚持了“相对性原理”。而且首先正确阐述相对论，认识到它是一个时空理论，并给出完整理论体系和几乎全部结论的也是爱因斯坦，而不是别人。所以说，爱因斯坦是相对论的唯一发现者。

事实上，在相对论发表之后，洛伦兹和庞加莱都曾反对它。洛伦兹后来接受了相对论，庞加莱则至死都未发表过赞同相对论的言论。

洛伦兹抱住绝对空间和以太概念不放，甚至主张放弃相对性原理。庞加莱虽然坚持相对性原理，主张放弃绝对空间，但他没有放弃“以太”。而承认“以太”实质上还是承认绝对空间的存在。

有一点需要解释一下。在相对论诞生之前，庞加莱于1900年在《时

间的度量》一文中曾经谈到：“光具有不变的速度，尤其是，光速在所有方向都是相同的。这是一个公设，没有这个公设，便不能试图度量光速。”这句话中“光具有不变的速度”，似乎是指“光速不变原理”。但从上下文看，庞加莱这句话是针对测量光速说的。众所周知，测量光速并不需要“光速不变原理”，但需要用“光速各向同性而且是一个常数”这一约定。他在这里强调的是同一个参考系中光速是点点均匀且各向同性的，即光速是一个常数 c 。而“光速不变原理”指的不是这一点，而是指光速在不同惯性系中相同。庞加莱从来没有在任何一个地方明确指出过“不同惯性系中的光速相同”。而且，承认“光速不变原理”就将直接导致“同时相对性”的概念，庞加莱也没有在任何地方谈到过“同时的相对性”。因此不能依据这句话，认为庞加莱在相对论发表之前就已认识到了“光速不变原理”。

1900年前后，庞加莱已是一位举世闻名的数学大师，爱因斯坦不过是一名初出茅庐的青年学者。庞加莱为相对论的诞生做了许多重要的基础性工作。他正确指出时间的测量依赖于对信号传播速度的约定。具体来说就是他认为“测量时间”需要首先“约定”（或者说“规定”）光速，他建议约定真空中的光速各向同性而且是一个常数。庞加莱正确地阐述了相对性原理，指出了洛伦兹理论的不足。一些学者认为相对论应是庞加莱与爱因斯坦共同创建的。

爱因斯坦与庞加莱只在学术会议上见过一次面。青年爱因斯坦当时非常渴望庞加莱支持相对论。那次会面回来后，爱因斯坦很沮丧，告诉他的朋友：“庞加莱根本不懂相对论。”事实上，庞加莱直到去世也未发表过赞同“相对论”的意见。

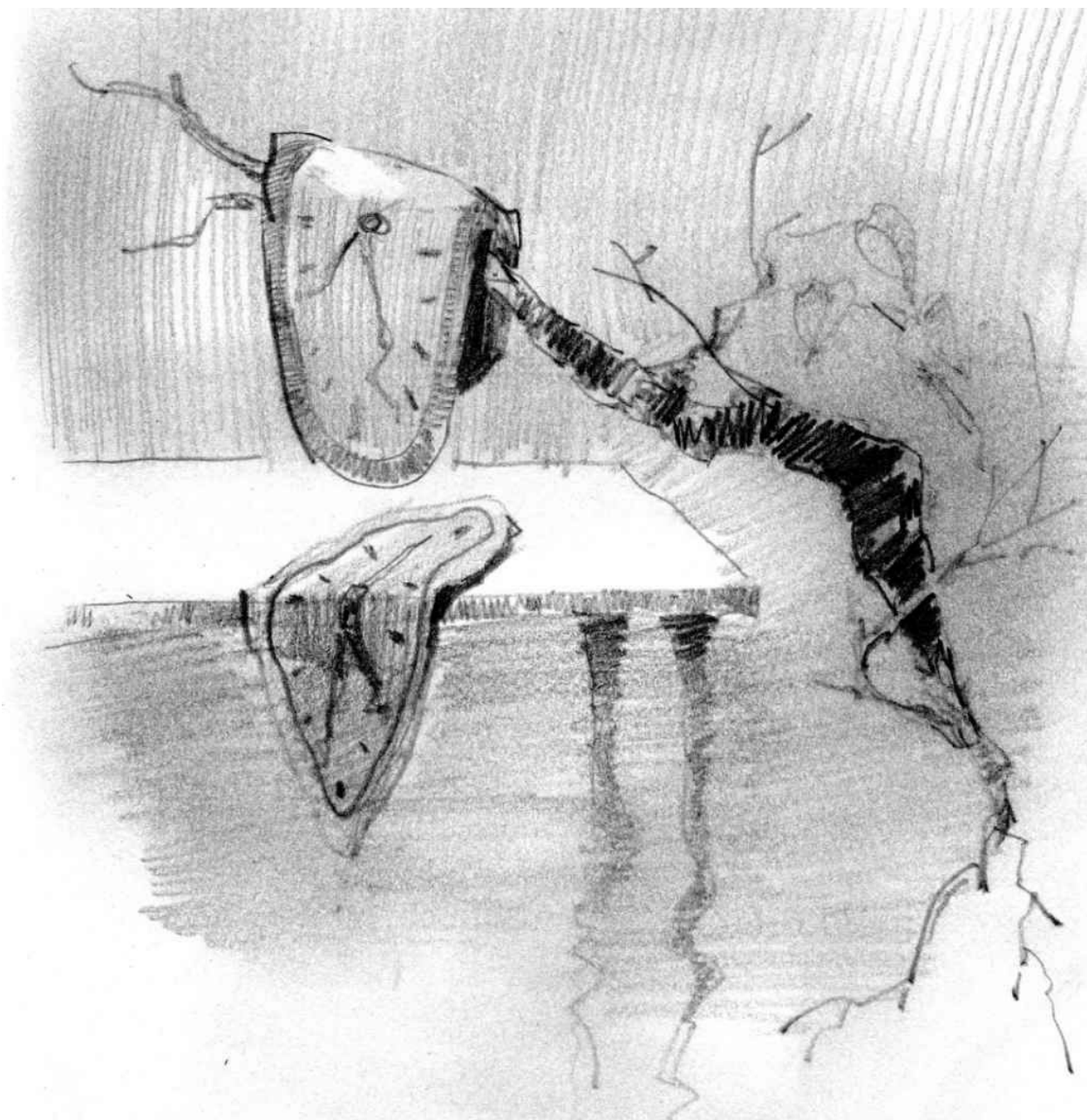
庞加莱对爱因斯坦的评价不十分高。他去世前不久，应苏黎世工业大学的邀请，对爱因斯坦申请教授位置发表了以下意见：“爱因斯坦先生是我所知道的最有创造思想的人物之一，尽管他还很年轻，但已经在当代第一流科学家中享有崇高的地位。……不过，我想说，并不是他的

所有期待都能在实验可能的时候经得住检验。相反，因为他在不同方向上摸索，我们应该想到他所走的路，大多数都是死胡同。不过，我们同时也应该希望，他所指出的方向中会有一个是正确的，这就足够了。”后来的研究表明，历史与这位数学大师开了一个极大的玩笑：爱因斯坦在1905年指出的所有方向都是正确的。

杨振宁教授指出，洛伦兹与庞加莱都曾非常接近相对论的发现。但是洛伦兹只有近距离的眼光，没有远距离的眼光，他只重视实验与观测，缺乏哲学思考；庞加莱只有远距离的眼光，缺乏近距离的眼光，他只重视数学和哲学思考，但忽视实验与观测。爱因斯坦既有近距离眼光，又有远距离眼光；既重视实验与观测，又重视哲学思考。最终，洛伦兹与庞加莱都没有发现相对论，只有爱因斯坦发现了它。

不过，爱因斯坦也承认许多人已经接近了狭义相对论的发现。他后来说：“如果我不发现狭义相对论，5年之内就会有人发现。”

第二讲 弯曲的时空——广义相对论



绘画：张京

这一讲介绍爱因斯坦一生最得意的成就——广义相对论。

首先要说明，相对论这个名字不是爱因斯坦起的，而是洛伦兹起的。因为洛伦兹在爱因斯坦之前就提出了洛伦兹变换，但他完全是依据绝对空间得到的。爱因斯坦的理论出来以后，得到的惯性系之间的坐标变换公式与洛伦兹变换相同，但是物理解释却很不一样。洛伦兹为了在辩论的时候分清楚我的理论和你的理论，就给爱因斯坦的理论取了个名字叫“相对论”，爱因斯坦觉得这名字还可以用，就接受了。以后“相对论”这个名字就留下来了。但当时所说的相对论指的是狭义相对论，就是我上一讲讲的那一部分内容，是讨论高速运动的物体会会有什么特点的理论。

1. 狭义相对论的困难

牛顿的力学完成的时候，物理学界都感到一片明朗，好像什么问题都搞清楚了，所以英国的一位诗人——波普，写了一首诗赞扬牛顿，说：

自然界与自然界的规律隐藏在黑暗中，

上帝说：“让牛顿去吧！”

于是一切成为光明。

可是相对论出来以后，这些感到光明的人大部分都感到糊涂，弄不懂相对论，当时能够听懂相对论的人是凤毛麟角。一般人都弄不懂，有的人就有疑问。但是大的物理学家们，一般不敢说得很难听。因为相对论出来以后，就有几位著名的物理学家说它是对的，比如德国的普朗克、能斯特、劳埃，法国的居里夫人和郎之万，还有英国的爱丁顿，这些都不是吃干饭的，都是很棒的物理学家，他们都说相对论正确，而且

予以很高的评价。所以那些自己觉得相对论不对的人，也不敢说得太难听，但是确实很怀疑。另一位诗人就把波普的诗给续了一段：

但不久，魔鬼说：“让爱因斯坦去吧。”

于是一切又回到黑暗中。

爱因斯坦的相对论到底有没有问题呢？当时的情况是这样，凡是觉得它有问题的人，说的那些问题其实都不是问题，都是自己没有弄懂相对论造成的。但是相对论真的有问题。爱因斯坦本人意识到了他的相对论有问题。

惯性系无法定义

那么爱因斯坦觉得他的相对论有什么问题呢？他的相对论建立在相对性原理和光速不变原理这两条原理的基础之上，它用到一个很基本的概念，就是惯性系。他知道自己理论建立在惯性系的基础上，可是现在惯性系却无法定义了。

为什么无法定义呢？牛顿认为存在一个绝对空间，牛顿说凡是相对于绝对空间静止或者作匀速直线运动的参考系就是惯性系。现在没有绝对空间了，那么惯性系就难以定义了。最初的一些人，甚至后来的很多人都认为，似乎可以用牛顿第一定律来定义惯性系：如果一个不受力的质点，在参考系中保持静止或者匀速直线运动的话，这个参考系就是一个惯性系。也就是说用牛顿第一定律来定义惯性系。很多人认为可以，其实不行。

爱因斯坦很快就意识到了这个定义不行。为什么不行呢？如果人家问你：你怎么知道这个质点没有受力呢？你要给“没受力”下个定义啊！有人可能会说，没有与其他物体挨着就是没受力啊！那不一定，像电磁力可以作用在带电或磁的物体上，但却是看不见的。最好的一个定义是说，在惯性系当中，一个质点保持静止或者匀速直线运动状态，它

就“不受力”。但是这种定义方式，定义“惯性系”要用到“不受力”这个概念，定义“不受力”又要用到“惯性系”这个概念，这是一个逻辑循环，所以这是不行的。爱因斯坦意识到惯性系的定义有了问题，这使相对论的基础变得可疑了，这个问题必须解决。

爱因斯坦反复思考惯性系如何定义，百思不得其解。有的人可能一辈子就琢磨这个定义了。但爱因斯坦的思路确实跟别人不一样。他开始想别的办法了。爱因斯坦想：惯性系既然不好定义，我就干脆不要惯性系了。我把相对性原理推广，推广到任意参考系。不是说物理规律在所有的惯性系当中都一样，而是说物理规律在所有的参考系中都一样。不要惯性系，定义惯性系的困难自然就不存在了。他这个想法很好。但是，不要惯性系，马上就有一个问题。惯性力怎么办？所有的非惯性系都存在惯性力，比如说一个转动的参考系，它有惯性离心力，有科里奥利力。一个加速系中的所有物体，都会受到反方向的惯性力。这些惯性力怎么处理？爱因斯坦觉得这仍然是个问题。

万有引力定律放不进相对论的框架

爱因斯坦注意到，自己的“相对论”还存在另外一个问题，就是万有引力定律写不进相对论的框架。当时只知道两种力，电磁力和万有引力，电磁学是跟相对论一致的，但万有引力却放不进相对论的框架，爱因斯坦觉得很遗憾。所以他也一直考虑这个问题，想把万有引力定律加进相对论的框架。努力了一段时间，但始终加不进去。

不过，爱因斯坦很快就注意到，万有引力和惯性力有相同的地方。什么地方呢？就是都与质量成正比。别的力不一定和质量成正比，只有万有引力和惯性力这两种力是跟质量成正比的。他觉得这两种力好像有点什么关系。他觉得自己所认为的相对论的两个困难，一个与惯性系有关，一个与万有引力有关。这两个困难莫非本质上是同一困难？于是他开始把这两个困难联系起来考虑。

当时已有人对惯性力的起源作过一些猜想。大家都知道，除去惯性

力以外所有的力都起源于相互作用，都有对应的反作用力。但是惯性力不起源于相互作用，它不满足牛顿第三定律，不存在对应的反作用力。

有绝对空间吗？

为什么会有惯性力呢？牛顿认为存在一个“绝对空间”，当一个物体相对于绝对空间加速的时候，就会受到惯性力，如果它不相对于绝对空间加速就不会受到惯性力。牛顿是在存在绝对空间的前提下来解释惯性力的。后来奥地利有一位物理学家马赫，这人是个三流的物理学家，他的贡献主要是“马赫数”，空气动力学中的“马赫数”，当然这也是贡献，能把名字留下来就算不简单。马赫虽然对物理学的具体贡献不是太大，但他对爱因斯坦有重大影响。你别看他三流的物理学家，他敢说祖师爷不对。他批判牛顿，认为牛顿说的绝对空间根本就不存在，所有的运动都是相对的。爱因斯坦在大学刚毕业时看过马赫的书，“呀，马赫讲得太对了，所有的运动都是相对的，根本就不存在绝对空间。”所以爱因斯坦不愿意放弃相对性原理。马赫的这一思想引导他建立了狭义相对论。

牛顿的水桶实验

牛顿当时为了论证绝对空间的存在，也为了论证惯性力起源于相对于绝对空间的加速，曾经提出过一个思想实验，叫水桶实验。如图2-1所示，他设想有一个桶，里面装有水。桶静止，水也静止的时候，水面是平的（见图2-1（a））；然后让水桶以角速度 ω 转起来，刚开始的时候由于桶壁的摩擦力小，水没有带动起来，桶转水不转，水面还是平的（图2-1（b））；然后水慢慢被带动起来了，跟桶一块转，这时候水面就成凹的了，这是第三种情况（图2-1（c））；第四种情况，桶突然停止，水还在转，这时候水面仍然保持凹形（图2-1（d））。为什么呢？牛顿说三、四这两种情况水受到了惯性离心力，而一、二这两种情况水没有受到惯性离心力。

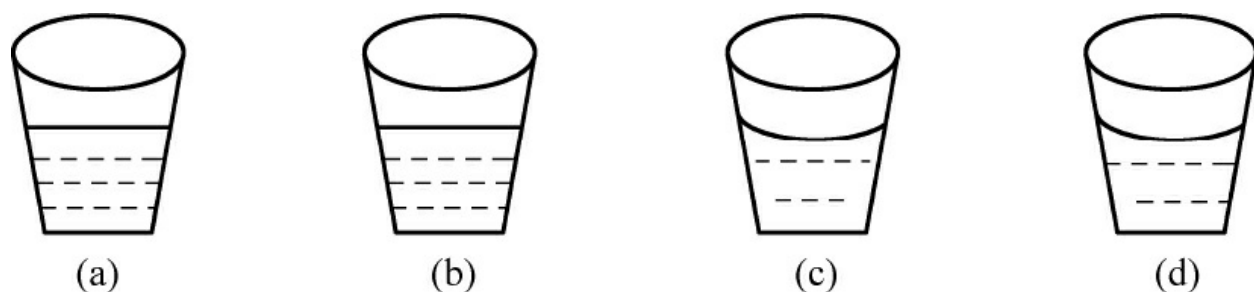


图2-1 水桶实验

一、三这两种情况水相对于桶都是静止的，但情况一水没有受到惯性离心力，情况三受到了惯性离心力。二、四这两种情况水相对于桶都是转动的，但情况二水未受到惯性离心力，情况四却受到了惯性离心力。看来受不受到惯性力跟水相对于桶的转动无关。

牛顿说，这个实验表明，受不受到惯性力跟水相对于桶的转动无关。那么跟什么有关呢？牛顿说，水桶实验表明存在一个绝对空间，只有相对于绝对空间的加速才是真加速，相对于绝对空间的转动才是真转动，才会受到惯性离心力。当时的物理学家都知道牛顿的水桶实验。牛顿用水桶实验论证了绝对空间的存在，同时也说明了转动是一种绝对运动。

马赫对牛顿的反驳

马赫出来说：“牛顿不对。”他就要面对这个实验。马赫反驳牛顿，说根本没有绝对空间，所有的运动都是相对的，那惯性力怎么起源的呢？马赫说水受不受到惯性力，跟水相对于桶的转动关系不大，他没说完完全没有关系，但是说关系不大。他认为，惯性力是宇宙中所有作相对加速的物质施加的作用。比如说水如果相对于宇宙中所有物质转动的话，就相当于水不动，宇宙中所有物质反着转，那么这些反向旋转的物质都对水施加一种作用，水就会受到惯性离心力。桶有没有影响？有，但桶的质量跟整个宇宙的质量相比是可以忽略的，所以桶对水的影响可以忽略。因此他认为，惯性力起源于相对于宇宙中所有物质的加速或者转动，起源于作相对加速运动的物质施加的作用。爱因斯坦看过马赫的

书后，觉得马赫讲得真是太对了。按照马赫的这个思想，惯性力也起源于相互作用，这种相互作用跟万有引力有某种类似，都与物质的成分和结构无关，只与它们的质量有关。马赫的思想加深了爱因斯坦的猜测：万有引力和惯性力之间可能有内在关系。

2. 等效原理

引力质量与惯性质量相等

这时候爱因斯坦进一步思考了一个问题，就是质量定义的问题。牛顿的《自然哲学之数学原理》是一部很完备的书，里面没有大的漏洞，逻辑关系非常严密。牛顿谈到了质量的定义，说“质量就是物质的量，质量等于体积和密度的乘积”，“质量正比于重量”，这些是他的原话。所以，“质量是物质的量”其实指的是物质的万有引力效应。牛顿在那本书的另外一个地方谈到质量是跟物体的惯性成正比的，那是跟牛顿第二定律有关的。牛顿意识到了用惯性效应来定义的质量和用引力效应来定义的质量可能不是一个东西。也就是说，质量有两种，一种是惯性质量 m_I ，另一种是引力质量 m_g 。这个 g 代表引力，牛顿的万有引力。

根据牛顿的判断，他觉得这两种质量不是一个东西。但实验表明呢，这两种质量可能是相等的。为什么呢？大家来看自由落体定律。自由落体运动，按照牛顿的理论，是在万有引力作用下的加速运动。万有引力可以用引力场强 g 乘上引力质量来得到：

$$F = G \frac{Mm_g}{r^2} = m_g g \quad (2.1)$$

牛顿第二定律是

$$F = m_I a \quad (2.2)$$

这两个等起来，就是

$$m_g g = m_I a \quad (2.3)$$

自由落体定律告诉我们，不管任何物体，随便什么物体，加速度 a 都是

等于 g 的，所有的物体不管质量，不管化学成分，它们的加速度都是一样的。如果 a 与 g 恒相等，那么 m_I 与 m_g 就相等。所以自由落体定律告诉我们，引力质量和惯性质量是相等的。

但是这个实验太粗糙了，于是牛顿又想用单摆实验来检验它。大家通常看到的单摆公式都是

$$T = 2\pi \sqrt{\frac{l}{g}} \quad (2.4)$$

实际上你们注意，在用微分方程推导的时候，质量 m 出现在方程的两边，一边代表引力质量，另一边代表惯性质量。只不过我们在学习的时候，在讲理论力学的时候，不区分这两种质量，于是就给消掉了。其实这两个质量定义不一样。牛顿注意到了这一点。如果你把这两个 m 保留的话，单摆周期公式就成为这样的

$$T = 2\pi \sqrt{\frac{m_I l}{m_g g}} \quad (2.5)$$

如果这两个质量对于不同的物体有差异， m_I/m_g 对各种物体不是同一个常数的话，单摆运动的周期，对不同物体就会有所不同。但牛顿没有观测到这种不同。他在千分之一的精度范围内证明了引力质量等于惯性质量。

爱因斯坦那个时代有个匈牙利物理学家Eötvös，中国人翻成厄阜，他用扭摆实验在 10^{-8} 的精度之内没有查到引力质量和惯性质量的差异。相对论发表以后，Dicke做到 10^{-11} ，俄罗斯的布拉金斯基做到 10^{-12} ，都严格地证明了引力质量和惯性质量相等。在爱因斯坦那个时代，精度最高的是Eötvös那个实验。爱因斯坦研究引力理论时知道这个实验。

等效原理：万有引力与惯性力等效

那时候爱因斯坦成天反复思考着引力与惯性力的问题。他当时还在专利局工作。有一天，他坐在办公桌旁，一边看蓝天白云，一边思考。

他突然想：假如有一人从楼上掉下来会是什么感觉呢？他想这个人可能是失重的感觉，没有重量。爱因斯坦后来说，这是我思想上的一次大的突破，这件事情引导我走向了广义相对论。

很快，爱因斯坦就提出了等效原理。这个原理是什么意思呢？就是万有引力和惯性力是等效的，是没法区分的。他说，如果有一个升降机（图2-2），外边是封闭的，里面的人看不见外面。升降机停在地球的表面上，里面的人具有重量。如果这个人拿着一个苹果，一松手这个苹果就落地。同样的，假如他处在远离所有星球的宇宙空间当中，是在一个火箭里面，虽然他没有受到重力，但是火箭在以加速度 a 加速，他也会同样感觉有重量，而且苹果会落地。也就是说，如果这个升降机是封闭的。他没法区分自己究竟是在一个有引力的星球表面上静止不动呢，还是在一个远离星球的地方作加速运动。再有一种情况，假如电梯的绳子断了，自由落体，在地球重力场当中自由下落，电梯里的人就会有失重的感觉。假如他在远离所有星球的地方作惯性运动的话，他是不是也会感受到失重？现在我们知道星际航行者就是这样的，他会感受到失重。他没法区分自己究竟是在引力场中自由下落呢，还是在不存在引力的空间中作惯性运动。因此引力场和惯性场是等效的，是不能区分的，这叫等效原理。

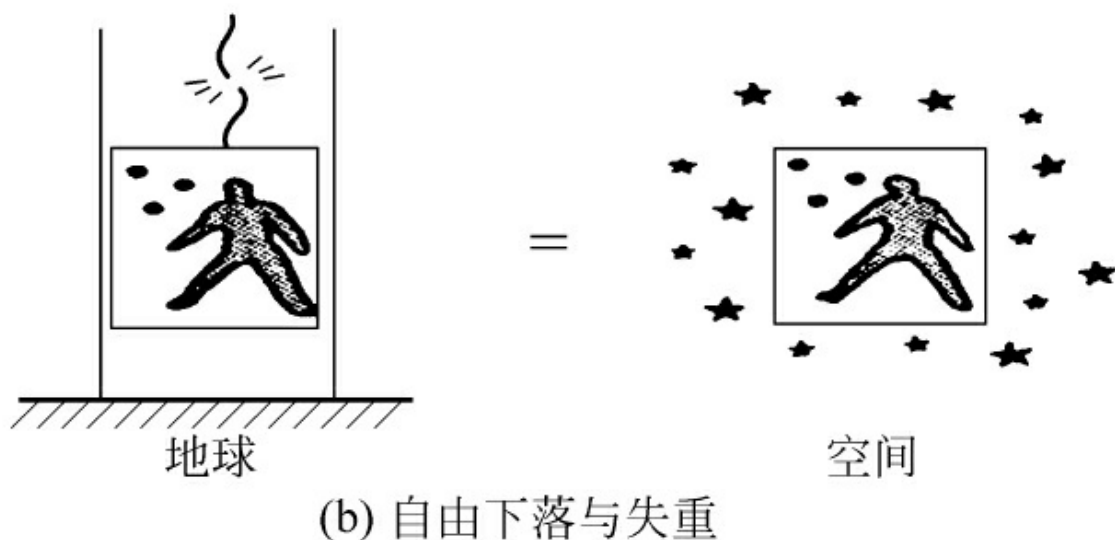
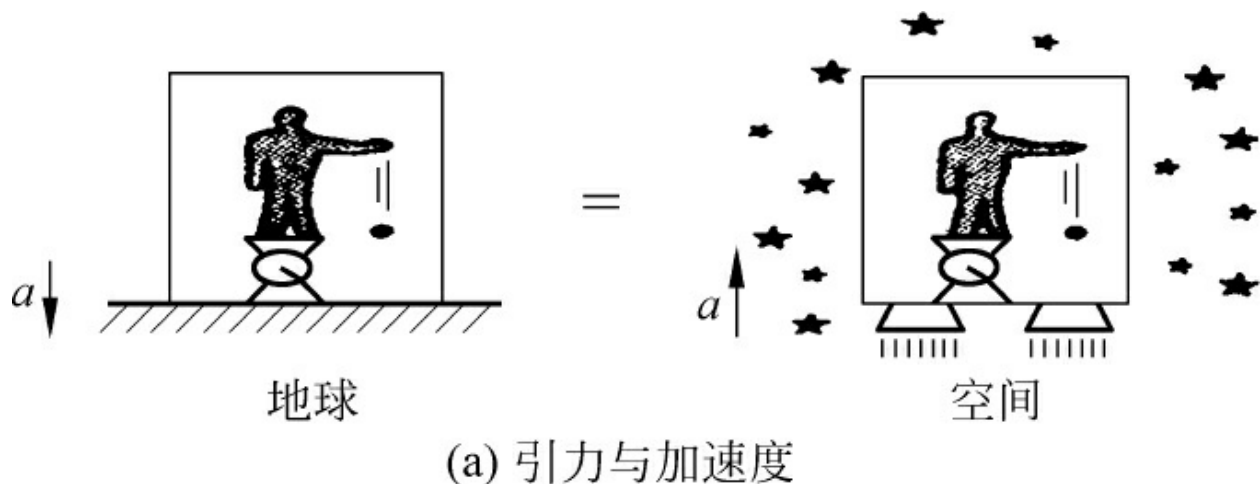


图2-2 爱因斯坦升降机

不过，引力场和惯性场的等效只在时空点的一点的邻域成立。只有在升降机无穷小的时候，引力场和惯性场才是不能区分的。假如升降机有一定大小，例如我们通常的电梯都有一定的大小，如果你在电梯地板的每一点都摆一个重力仪的话，你就会感觉到力线有一个向地心的汇聚效应。而你要在星际航行的火箭上摆上重力仪的话，力线就是平行的，所以在空间不是无穷小的情况下，还是能区分引力场和惯性力场的，不能区分的就是一点的邻域。这是学习等效原理最应该注意的一点。等效原理还分弱等效原理、强等效原理，由于时间关系我们就不说了。

思想的飞跃：引力可能是几何效应

爱因斯坦到这个时候，物理上的思考已经开始有了眉目。他觉得：第一，为了克服惯性系的困难，可以把相对性原理推广为广义相对性原理，就是说不用惯性系了，认为在所有的参考系中物理规律都一样。不过这时候会出现惯性力的困难。此时，他认识到了惯性力的困难跟万有引力的困难可能是同一个困难。而且这个时候他思想产生了一次重大的飞跃，就是认为万有引力可能不是真正的力，而是一种几何效应。他为什么会这么想？你们看，自由落体定律，任何物体不管质量、化学成分和物质结构，下落规律都一样。如果只看自由落体定律，还看不清楚的话，你还可以考虑斜抛物体。在真空当中以某一个角度斜抛一个物体，不管是个金球、铁球还是个木头球，如果抛射角度保持不变，球脱离弹射器时的初速也保持不变，那么它们描出的轨迹就全一样。跟它们的成分、质量、物理性质和化学性质都毫无关系。这跟所有的物理定律都不一样，一般物理定律和化学定律都是跟物质的成分、结构、质量等有关的。但自由落体和斜抛物体是完全在单纯的万有引力作用下按照牛顿定律的运动，这类运动的规律跟物质的成分和质量都没有关系。这时爱因斯坦突然想到这类运动会不会是一种几何效应，因为几何效应肯定与物体的质量、成分无关。这是非常非常大胆的、思想上的飞跃。

所有的创新性的发现都不是靠着逻辑推理推出来的。逻辑推理推出的只会是已有结论的另一种表现形式，或者一些特殊情况下的例子。所有真正的科学发现都是猜出来的，然后用实验去验证。

3. 神奇的黎曼几何

爱因斯坦猜测万有引力可能是时空弯曲的表现，那么这时候他就要用到弯曲时空当中的几何学了。于是爱因斯坦找他的同学格罗斯曼帮忙。格罗斯曼留在苏黎世工业大学以后，主要搞数学，当时已是数学物理系主任。他查了一些资料后告诉爱因斯坦，现在有些意大利人正在研究黎曼几何，可能这个东西对你有用。后来，格罗斯曼也参加了爱因斯

坦的研究，所以爱因斯坦探索广义相对论的早期论文有些是跟格罗斯曼合作的。

那时候已经有了黎曼几何，其实黎曼几何对爱因斯坦不是完全生疏的。爱因斯坦在专利局工作期间，与几个年轻人自发组织了一个读书俱乐部，他们取了个名字叫“奥林匹亚科学院”，就那么三四个人，有学物理的，有学工程的，还有学哲学的。大家在一起读一些科学、哲学或其他数学方面的书，边读边议论。他们经常在爱因斯坦的家里面读书。爱因斯坦的夫人米列娃常常坐在那儿，但是她一般不发言，只是静静地听他们在那里讨论。他们当时看过马赫的《力学史评》，还看过庞加莱的《科学与假设》，庞加莱在这本书里用科普的方式提到了一点黎曼几何。

“平行公理”导致的疑难

为了让大家更清楚地了解弯曲时空中的几何学，我们简单说几句黎曼几何的创建。我们先说欧几里得几何。公元前300多年，埃及被希腊人占领，当时埃及的国王是希腊人，姓托勒密，不是搞地心说的那个托勒密，只是同一个姓。托勒密一世和二世国王非常喜欢科学和建设，他们建了一个亚历山大科学院，就在埃及北部的海港城市亚历山大。还设立了科学基金资助科学家们进行研究。欧几里得就在那个地方工作，欧几里得把古埃及人研究大地测量、研究尼罗河泛滥后平分土地、修建金字塔等积累的几何知识总结成一本书，就是介绍欧几里得几何的名著《几何原本》。欧几里得之前的人的贡献已经不清楚了。他那个时候集其大成。

欧几里得几何里面有很多公设，就是公理。从这些公理可以推出所有的定理和推论。其中有个第五公设，就是平行公理。我们大家都知道，就是过直线外一点能引一条并且只能引一条直线跟原直线平行。很多人觉得这个公理有点长，是不是可以从其他的公理推出来。于是就有许多数学家在那儿推导，推了一千多年、两千年的样子，所有的人都推

不出来。或者有时候高兴一下，说：“哎呦，推出来了，你只要假设三角形三内角之和是 180° 就可以推出来。”但是，这是一个同等的假设，你假设三角形三内角和为 180° ，跟你假设平行公理是一个事情，反正你得假设一个，所以一直证不出来。很多人为此耗费了自己的毕生精力。

鲍耶与高斯的探索

最早对平行公理的研究做出突破性贡献的人之一，是匈牙利年轻数学家鲍耶。鲍耶当时在一个数学系学习，他父亲是高斯的同学。鲍耶在证明平行公理的时候使用了反证法，他想，假如过直线外的一点可以引两条以上的直线跟它平行，那会怎么样呢？原本他想推出错误来，结果总也推不出来。他突然产生一个思想飞跃，这个飞跃非常重要。他考虑是不是可以建立另外一种几何，假定过直线外的一点可以引两条以上的平行线，这样就能建立一套完备的新几何。他把自己的想法告诉了父亲，他父亲一听儿子在研究这个，简直难过极了，说：“我的儿子，你可千万别干这个了，你爹我就是因为研究这个，最后几乎一辈子一事无成啊！你可千万别走这条路了。”后来，他爹仔细一看鲍耶写的东西，觉得儿子的研究还真有点道理，于是挺高兴，就把这些东西写信告诉高斯。高斯看了以后说：“我实在没法赞扬你的儿子，因为赞扬他就等于赞扬我自己，其实你儿子的想法我前些年就有了。”鲍耶听了以后非常生气，觉得高斯是想用自己的名望来篡夺他的研究成果，一气之下不干了。他父亲最后把儿子的成果作为附录，附在自己出版的一本数学教科书后面。由此世人才知道鲍耶做出过重大贡献。

罗巴切夫斯基的奋斗

不过最早提出并建立完整的新几何的人，不是鲍耶而是俄罗斯喀山大学的教授罗巴切夫斯基。他也在研究平行公理，也是用反证法，最后他也想到，会不会过直线外一点可以引一条以上的平行线，那样的话是不是可以得到一种新几何。他就把论文寄给彼得堡科学院，彼得堡科学院的院士们一看，这个教授简直是稀里糊涂，过直线外一点怎么可以引

两条平行线啊！这不是胡说八道嘛！不久，罗巴切夫斯基又来信了，又发来论文了。科学院的人说这个教授怎么回事啊，这点事情都不明白还当教授？然后彼得堡科学院的几个数学家就做了一个决议，说是以后凡是罗巴切夫斯基先生有关这方面的论文，我们都可以不必审稿了，肯定不要。罗巴切夫斯基只好把论文发表在喀山大学学报上，这些论文比鲍耶的工作还要早两年。由于在国内得不到支持，罗巴切夫斯基后来就到欧洲去周游交流，看看大家的反应怎么样。结果没有一个人表态支持他。他到德国发表了演讲，高斯听了演讲，没有说什么，当时高斯已经很老了。高斯只是建议德国科学院授予他通讯院士的称号，但是没有提他创建新几何的事。高斯在自己的日记和给朋友的信中说：“我相信，当时在会场上只有我一个人听懂了罗巴切夫斯基先生在讲什么内容。”但是高斯不敢表态。为什么呢？因为欧几里得几何是教会支持的，哥白尼、布鲁诺他们的前车之鉴，使高斯顾虑很大，他想：我还是少插这一脚吧，反正我干的事多了，也不在乎这点东西。所以他不掺和。高斯去世以后，这些东西才披露出来。罗巴切夫斯基从欧洲回去以后，因为德国人也承认了他的学术水平，后来当上了喀山大学的校长，继续研究新几何。但是俄罗斯国内还是没有人承认。罗巴切夫斯基晚年双目失明，最后靠着口述，他的学生记载，把他的新几何搞了出来，这就是罗氏几何。

黎曼集其大成

过了些年以后，又有一个年轻的数学家黎曼，提出：过直线外一点一条平行线也引不出来，以这条公理为基础建立起另一套几何，这就是黎氏几何。黎曼又把欧式几何、罗氏几何、黎氏几何综合起来统一成黎曼几何。他用上述工作在哥廷根大学做了一个求职报告，争取一个讲师位置。由此可以想见，哥廷根大学的数学水平有多高了！

什么是黎曼几何？

实际上，黎氏几何是一种正曲率空间的几何，在二维情况下，就是

球面几何（图2-3（a））；罗氏几何是一种负曲率空间的几何（图2-3（b）），在二维情况下，就是伪球面和马鞍面上的几何；而欧几里得几何是一种零曲率空间的几何，在二维情况下，就是平面上的几何。它们描述不同曲率的空间，如表2-1所列，三种几何都对。

表2-1 三种几何的对比

	空间曲率	平行线	三角形三内角之和	圆周率	例
黎氏几何	正	无	$> 180^\circ$	$< \pi$	球面
欧氏几何	零	一条	$= 180^\circ$	$= \pi$	平面
罗氏几何	负	两条以上	$< 180^\circ$	$> \pi$	伪球面

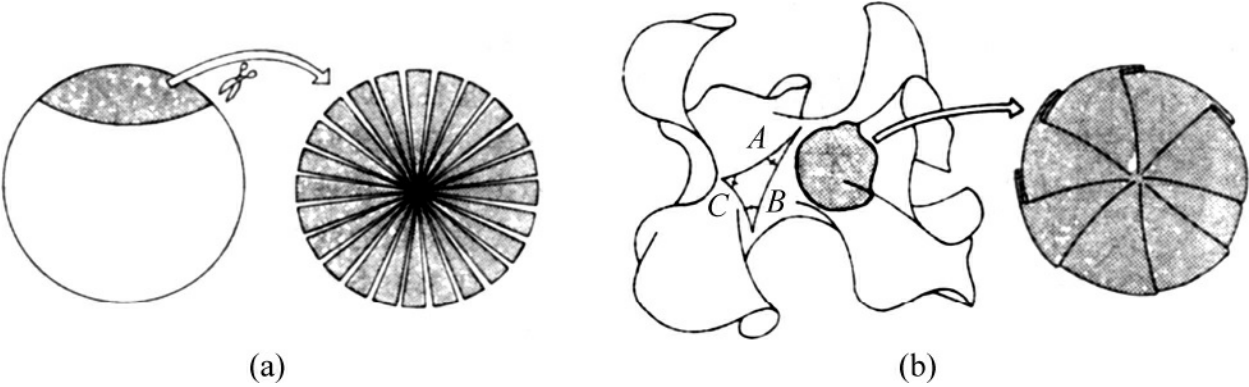


图2-3 正负曲率的空间

那么在弯曲空间当中，怎么定义直线呢？显然没有直线！但有短程线。所谓短程线就是两点之间最短的线。因为伪球面和马鞍面大家不那么熟悉，我们以球面几何为例来说明弯曲空间中的几何（图2-4）。球面上的短程线就是大圆周，你用球表面上的两点和球心这三点作一个平面，截出来的那个圆周——大圆周，就是短程线。比如说，赤道是短程线，所有的经线都是短程线，但是除去赤道外所有的纬线都不是短程线，因为它们都不是大圆周（图2-4）。地球表面两点之间最短的距离是沿大圆周的，所以从中国飞往美国和加拿大的飞机，并不是直接向东横越太平洋走的。它是从北京起飞以后往东北方向飞，穿过俄罗斯的西伯利亚，一直飞到白令海峡的北边，贴着阿拉斯加的北部沿海飞过去，

再进入加拿大，进入美国。有人说：“这不是绕了一个大弯吗？”不是绕了个大弯，那是真正最近的路线。黎氏几何，过直线外的一点引不出一条平行线是说什么呢？是说在一个大圆周之外，你不能再作一个大圆周跟它不相交。对不对？你想赤道是个大圆周，你能在赤道外再作一个大圆周跟它不相交吗？根本不可能。另外呢，在黎氏几何中三角形三内角之和是大于 180° 的，因为时间问题，今天我们就不讲了。

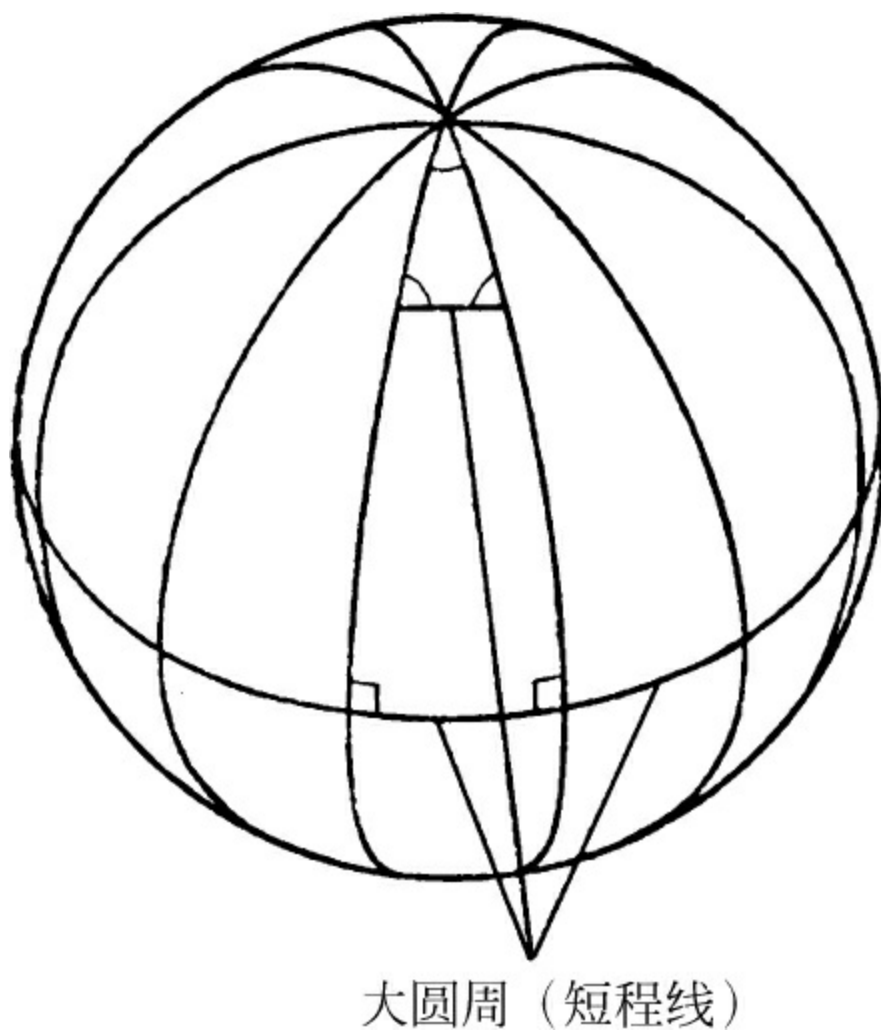


图2-4 球面上的大圆周和三角形的三内角

4. 广义相对论的创建

爱因斯坦场方程

在爱因斯坦的时代，黎曼几何已经有了。爱因斯坦在格罗斯曼的帮助下熟悉了黎曼几何，但是刚开始摸索的时候并没有得到正确的方程。后来他到了德国，与希尔伯特进行了几次讨论以后，终于找到了正确的方程。这就是广义相对论的基本方程——爱因斯坦方程，或叫场方程，

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu} \quad (2.6)$$

左边是时空曲率，右边是物质的能量动量，常数 κ 实际上是 $8\pi G/c^4$ ， G 就是万有引力常数， c 是真空中光速。你们看着简单，实际上它是二阶非线性偏微分方程组，10个二阶非线性偏微分方程组成的方程组，左边表示时空弯曲，右边表示物质的存在。这就是广义相对论的最基本的方程。这个方程解起来很困难，谁如果能求出来一个解，就可以以他的名字命名。到目前为止，有用的解没有几个，大部分解虽然数学上正确，但是物理上找不到对应，物理学家兴趣不大。因为物理学是一门实验和测量的科学。

万有引力不是力

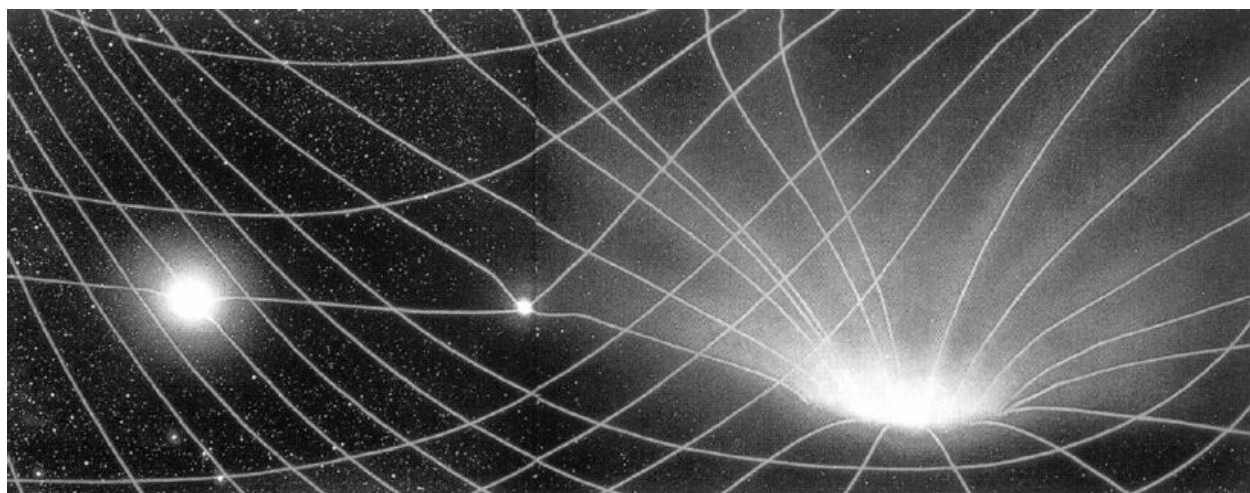
好，我现在就来定性地解释一下弯曲的时空。举个例子，我拿着一个粉笔头，一松手它就掉下来了，按照牛顿第二定律和万有引力定律，这是一个在万有引力作用下的匀加速直线运动。按照爱因斯坦的广义相对论，万有引力根本就不是什么力，只是时空弯曲的表现，松手之前你用了力拽着粉笔头，一松手这个粉笔头就没有受到力了，就自由下落，它作的是惯性运动。

再看行星绕日的运动，行星绕日的运动可以用万有引力定律和牛顿第二定律联立起来，严格地计算出它的椭圆轨道。现在我们发射人造卫星，也全部用的是牛顿力学，因为牛顿力学计算起来简单，而且在太阳引力场中足够精确。如果你用广义相对论的方程算，那就复杂多了。用牛顿第二定律和万有引力定律的联立，我们能够准确地预报卫星在几点几分过什么地方，非常精确地预报。按照牛顿力学，太阳用万有引力吸

引着地球，使地球依照牛顿第二定律，围绕着它转，走一个椭圆轨道，这是一种变加速运动。但是按照爱因斯坦的广义相对论，这是惯性运动，因为万有引力不是力，行星没有受任何力，绕着太阳转动是一种惯性运动，没有受到任何力的自由运动。

如何理解弯曲时空

图2-5是一个示意图，一颗恒星把周围的空间压弯了，这不是什么真正的物理图。我们可以打个比方，比如说四个人拽开一张床单，床单是平的，小玻璃球搁在上面不动，你一滚它就作匀速直线运动。但是如果床单中间放上一个铅球，就把床单压弯了。再将玻璃球搁在上面，玻璃球动不动？不动行不行？不行。它会滚到铅球那里去。我们可以把铅球想象成地球，这个玻璃球想象成粉笔头，它滚过去了。按照牛顿式的解释，就是那铅球（地球）用万有引力吸引这个玻璃球（粉笔头）；而按照爱因斯坦式的解释，铅球（地球）使周围的空间弯了，在弯曲的空间当中，这个玻璃球（粉笔头）就自然地滚过去了。同样的，你可以把那个铅球看作太阳，把玻璃球看作地球，你横着一扔，它就转起来了。这个玻璃球（地球）为什么不跑掉呢？也就是说行星为什么不逃离太阳呢？按照牛顿力学的解释，就是这个铅球（太阳）用万有引力吸引着玻璃球（地球），它跑不了；按照爱因斯坦的解释，就是铅球（太阳）让周围的空间弯了，在弯曲空间中，玻璃球（地球）作自由运动。也就是说，太阳让周围的时空弯了，在弯曲时空中，地球作自由运动，这个自由运动就是围着太阳转，所以它跑不了。这是一种直观的比喻。爱因斯坦认为，万有引力不是真正的力，而是时空弯曲的表现。行星绕日的运动是弯曲时空中的惯性运动。



1图2-5 弯曲的时空

伽利略的“错误”

我们这里顺便谈一下伽利略的一个错误，很多人认为他犯的一个错误。伽利略当时在谈论惯性运动的时候，对这个概念进行过解释。伽利略认为什么是惯性运动呢？他说静止或者匀速直线运动是惯性运动；再有呢，匀速圆周运动也是惯性运动。这后一个说法长期被认为是伽利略的一个错误。

我们今天知道匀速圆周运动确实不是惯性运动，但是伽利略为什么要说匀速圆周运动是惯性运动？最重要的就是那时候，不管是日心说还是地心说，都认为行星在作圆周运动。比如日心说，太阳在中心，地球围绕它旋转，这是一个圆周运动。地球为什么一直转，它又没有受到力？伽利略猜想这可能是一种惯性运动。今天看来伽利略的猜测本质上还是对的，在他的潜意识里，意识到了行星绕日的运动可能是惯性运动。

弯曲时空中的直线——短程线

我们都知道，在地球表面上一个不受力的物体作的惯性运动是匀速直线运动，走的是直线。那么在弯曲空间当中，它走的是什么呢？走的是所谓短程线，就是两点之间最短的距离。不过呢，广义相对论的情况

实际上要复杂，因为它用的是“伪”黎曼几何，就是时间的一项跟其他的几项的正负号不一样，所以行星绕日走的这些轨道反而是两点之间最长的那条，但是习惯上都叫短程线，或者叫测地线，实际上是最长的一条。可能有人以为行星绕太阳转动的椭圆轨道就是短程线。不是！因为相对论所说的短程线是四维时空当中的曲线（即世界线），不是三维纯空间中的曲线。我们可以假定太阳在三维空间中不动，是一个点。但是在四维时空中它会描出一根与时间轴平行的直线来，那么行星绕日走的就是螺旋线。在四维时空当中，这根螺旋线是短程线，是作惯性运动的行星描出的世界线。如图2-6所示。

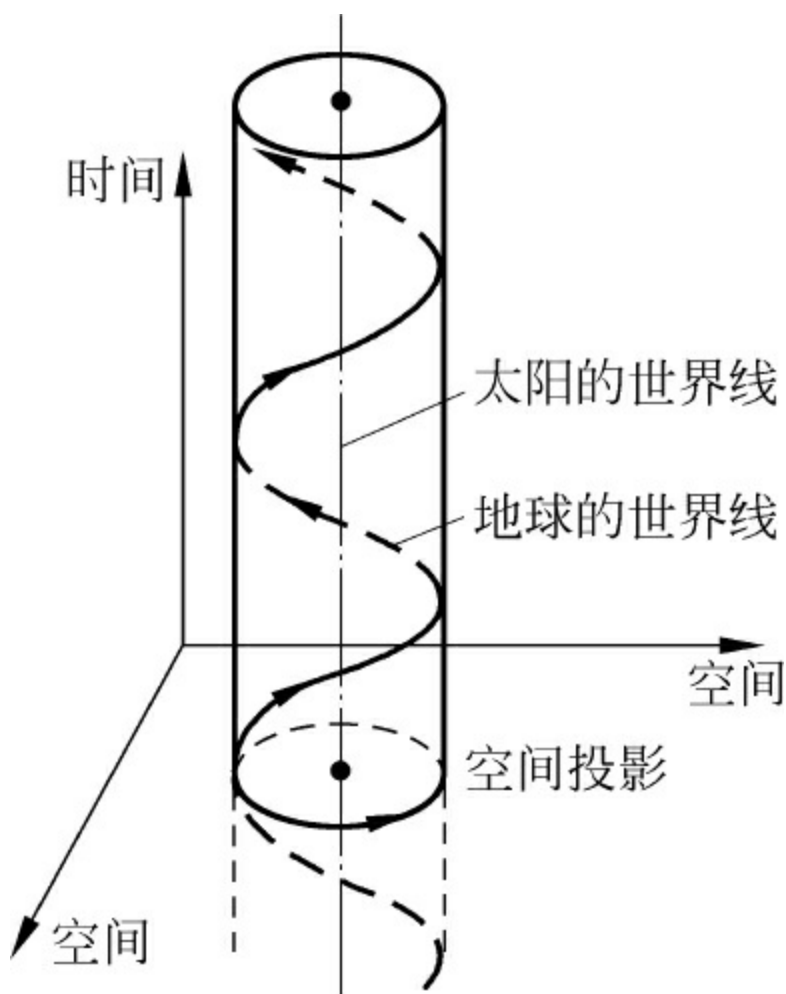


图2-6 太阳和行星在四维时空中的运动曲线

5. 广义相对论的实验验证

现在我们来讲一讲广义相对论的实验验证。爱因斯坦当年提出广义相对论的时候就说：“有三个实验可以证明我的广义相对论是正确的。”

时钟变慢与引力红移

第一呢，他说，由于时空弯曲，钟会走得比较慢，以前在狭义相对论中他说过运动的钟会变慢，现在他又说还有一个新效应，就是时空弯曲的地方钟也会变慢。时空弯曲得越厉害，钟走得越慢。因此太阳表面的钟会比我们地球表面的钟走得慢。怎么知道呢？可以在太阳表面放一个钟，然后去看一下。但实际上你没法放那个钟。就是放了个钟你也不敢去看啊，是不是？这怎么办呢？爱因斯坦说没关系。其实呢，太阳表面本来就有钟。他说：你看啊，每一种原子的光谱都有确定的光谱线。我们现在做化学分析不就是根据这些谱线的波长来判定化学元素吗？他说：“每一种原子都有特定的光谱线，每一根光谱线就表示，在这个原子当中有一个以这种频率在震荡着的钟。”你看看，他这想法跟一般人多不一样，恐怕很多人都想不到这一点。他说：“太阳表面有很多的氢，我们地球实验室也有氢。我们可以把太阳光中的氢光谱，跟地球实验室的氢光谱来比较，你就会发现由于太阳处的钟变慢，太阳上的氢光谱的振动的频率也要变慢。所以，太阳上的氢光谱的所有的光谱线会向红端移动，频率减小，波长增大。”大家把拍的照片一对比，果然如此。不过呢，用牛顿的万有引力定律也能算出光谱线的红移。为什么呢？一个光子从太阳那里跑过来，它要克服引力势能啊，对不对？光子要减少动能， E 等于 $h\nu$ ， E 一旦减小 ν 也会减小。所以从牛顿力学来看，光谱线也会出现红移。但是，对光谱线红移的牛顿力学解释和相对论解释，在高阶近似上是有差异的。

水星轨道近日点的进动

现在我们来谈另外两件事情。一件事是水星近日点的进动（图2-

7)。牛顿力学算出来的行星绕日运动都是一些封闭的椭圆。但实际上我们看到的行星绕日运动都不是封闭的椭圆，它会进动，这样一圈一圈地转起来，近日点在不断地前移。远日点也在移，但是远日点在天文观测上不好确定，近日点的移动，容易确定。离太阳越近的行星轨道进动越厉害。所以大家总以水星为例来讨论行星轨道近日点的进动。

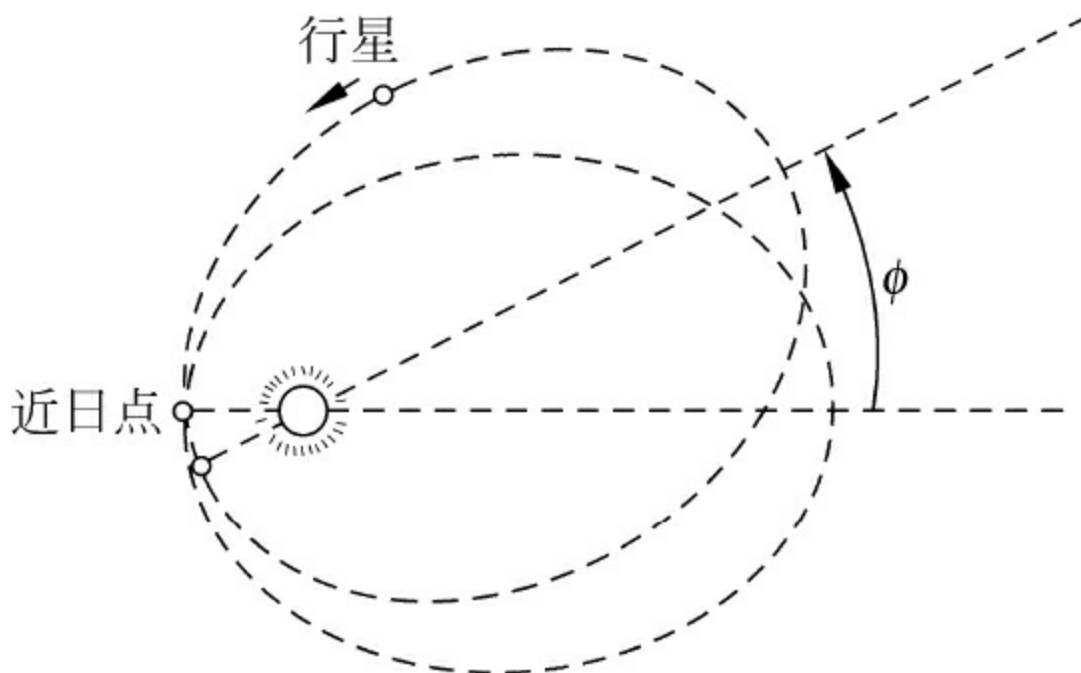


图2-7 水星轨道近日点的进动

在爱因斯坦那个时代已经知道，水星绕太阳的轨道不是一个封闭的椭圆，人们用天文学上的“岁差”和其他行星的影响来解释。当时观测到水星轨道每一百年有5600弧秒的进动，把所有可能的影响因素都考虑进去以后，有5557.62弧秒的进动能得到解释，但还有约43弧秒的进动没法解释。你看天文观测多精确，每一百年还有43弧秒的进动得不到解释！最先对这个问题进行研究的是勒维叶，就是预言海王星，发现海王星的那个人。

海王星是法国的勒维叶和英国的亚当斯首先预言，然后发现的。人们早就知道，除了地球以外还有五颗行星，就是金、木、水、火、土五

颗，这些是我们肉眼能看到的。天文望远镜出现以后又发现了天王星。后来发现天王星轨道的计算值跟实际观测值有偏离。当时英国的亚当斯和法国的勒维叶都怀疑是不是有一颗比天王星离太阳更远一点的行星对它有影响。亚当斯首先算出了结果，寄给了英国的格林尼治天文台，指出了这颗星的位置。格林尼治天文台的人一看，说：亚当斯？亚当斯是谁呀？从来没听说过。就搁一边了，没管。勒维叶独立算出来以后，就在科学杂志上发表了自己的结果。英国天文台的人看到勒维叶的文章后一想，咱们那个亚当斯不是也算过吗？就把亚当斯的东西拿出来看。准备观测一下，看是不是真的有一颗新的行星。正在英国人磨磨蹭蹭准备的时候，勒维叶急于知道自己的预言对不对，就把他的论文寄给了柏林天文台，因为法国的天文台不如德国的好。柏林天文台台长拿到这封信后，立刻叫他的部下当天晚上就去观测，结果一下就发现了海王星。这个报道一出来，英国人赶紧也去找，根据亚当斯的结论去找，也发现了海王星。所以海王星的发现，是万有引力定律的一个伟大成就。因为它用万有引力定律预言的，首先准确地预言了它，把它算出来的，然后由观测证实。

勒维叶取得这个成就以后就想，水星轨道的进动，是不是因为有一颗比水星离太阳更近的行星造成的？他就反过来算出了这颗星的位置。他很高兴又预言了一颗新行星。他给这颗新行星取名火神星，为什么叫火神星呢？因为离太阳近，温度高。于是有些人去观测，但怎么也找不到。一次，有个人真的看到了太阳表面上有个黑点在那儿动。以为发现了“火神星”，高兴了一场，后来发现那颗“火神星”不过是个太阳黑子，不是行星。所以水星这个43弧秒的进动问题就遗留下来了。

几十年以后，爱因斯坦的广义相对论算出水星轨道正好就有43弧秒的进动。就是说，不考虑别的因素，水星绕太阳转动的轨道就不是封闭的椭圆，一百年就有43弧秒的进动，水星轨道就不停地前移。爱因斯坦算出这一结果后特别高兴，因为他事先知道有这个进动，他希望他的新

理论能够解释这个进动，那么就比牛顿的理论优越了。在给洛伦兹和其他朋友的信当中，爱因斯坦说：“我的新理论算出了水星轨道近日点的进动，我高兴极了。你们知道我有多高兴吗？我一连几个星期都高兴得不知道怎么样才好。”水星轨道近日点进动是支持他的广义相对论的最重要的实验。因为它在二级近似上得到了精确的结果。

光线偏折

还有一个检验爱因斯坦理论的实验——光线偏折。按照广义相对论，太阳的存在会造成时空弯曲。图2-8中显示，在地球上的人看来，一颗恒星射来的光，如果没有太阳存在，走的是直线，光来自图中黑星所示的它的真实位置。有太阳呢，时空弯曲会使星的图像出现在图中白星所示的表观位置。这叫光线偏折。不过根据万有引力定律也能算出光线偏折，为什么呢？一个光子路过太阳附近，在重力场中下落，它应该走一条抛物线，所以光子的路径也应该弯曲，但两种解释的偏转角不一样，相对论预言的偏转角是牛顿理论预言的偏转角的两倍。

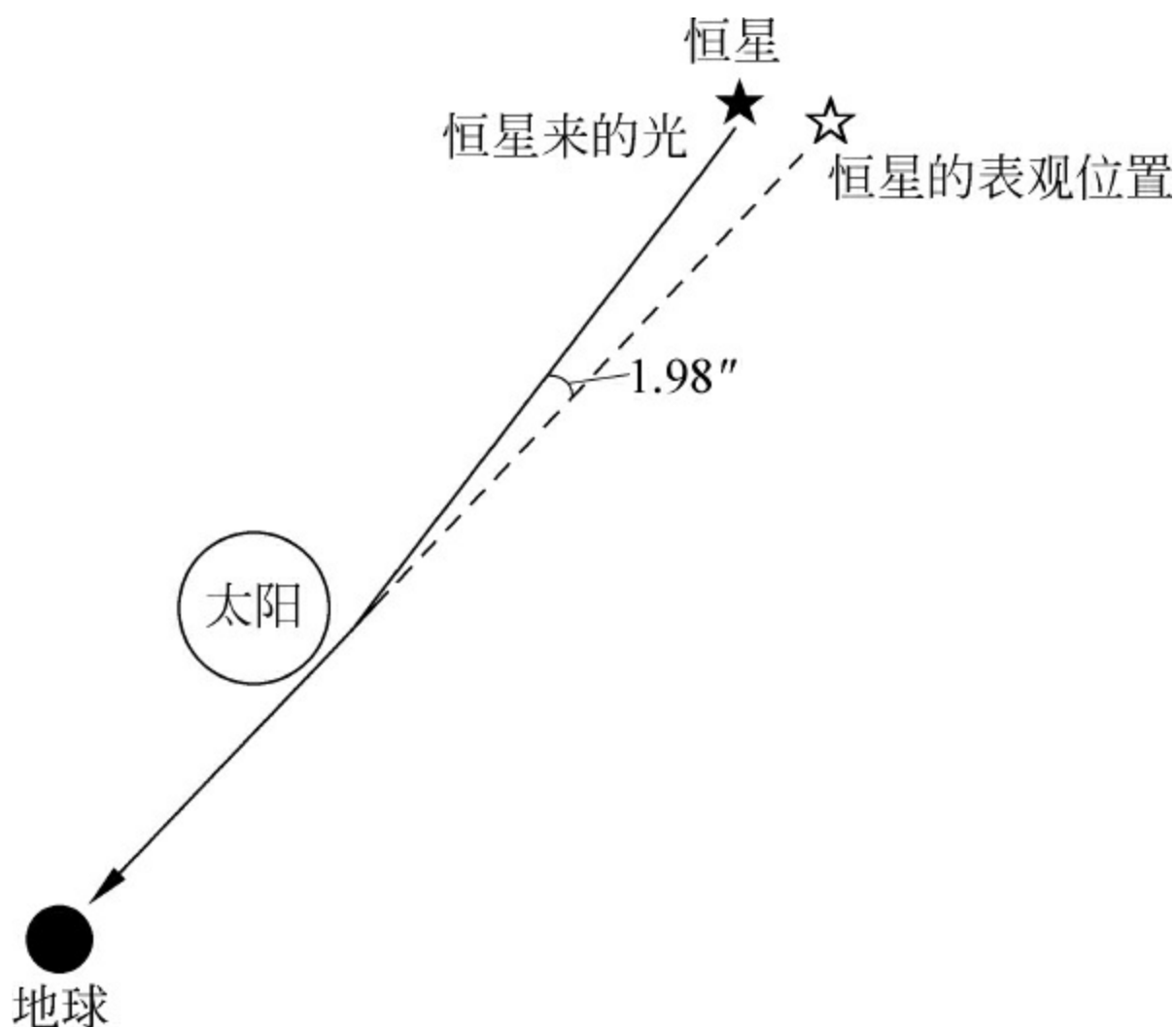


图2-8 光线偏折

在1915年发表广义相对论时，“一战”正在进行，战争中，英国、法国和德国都死了很多人，民族仇恨非常严重。为了减弱民族仇恨，增进英德人民之间的友谊，战后英国人拿出了一笔钱，说是要资助一些项目，这些项目要能够增进英德两国人民之间的友谊。爱丁顿教授就出来说要这笔钱，申请这笔钱检验爱因斯坦的广义相对论。他说广义相对论是德国的爱因斯坦提出来的，现在由我们英国人来检验，这不就能够增进两国人民的友谊吗？他拿到了这笔经费。

不过观测太阳附近的光线偏折有点困难，有太阳的时候，你怎么拍这个照片，大白天的，那么亮，根本拍不了。有没有办法能拍照呢？日

全食！是吧？当时，1919年恰好有日全食，爱丁顿带了一个观测组到西非的普林西比，还有一个助手带了另一个组到巴西。爱丁顿在西非的普林西比，遇到阴天，你看倒霉不倒霉，日全食的时候是阴天。不过就在日全食结束前几分钟，来了一阵风把乌云刮开了。哈呀！真是高兴极了，他们立刻抓住这几分钟的时间连续拍了15张照片。

巴西那个组倒是遇上了好天气，顺利地拍下了照片。但拍得的照片却令他们失望。日全食时那里是艳阳天，按理说应该得到很好的观测结果，但是因为天气太好，阳光把仪器晒得太热了，以致照片发生了变形。不过，他们把照片形变造成的误差排除后，也得到了可用的结果。

没有太阳的情况怎么拍呢？是这样的，地球绕着太阳转，日全食的时候太阳背后的星空，半年之后地球转到太阳的另一边时，就在夜间出现。所以你几个月后就可以在夜间拍照片，拍那个不存在太阳的星空。两组照片一比较，就可以得出偏转角了。

结果出来了，两个小组测得的偏转角与广义相对论都符合得很好，都得到了与广义相对论预言一致的结果。爱丁顿在报告中说：“根据牛顿的理论，偏转角是0.875弧秒；根据爱因斯坦的理论，偏转角是1.75弧秒。两个组观测到的偏转角分别是1.61弧秒和1.98弧秒，实验观测支持了爱因斯坦的广义相对论。”有人问爱因斯坦：“您有什么感想？”爱因斯坦说：“我从来没想过会是别的结果。”他非常自信。

广义相对论是爱因斯坦最得意的成就，他说：“狭义相对论如果我不发现，五年之内就会有人发现。”因为很多人都接近狭义相对论的发现了。“广义相对论如果我不发现，50年之内也不会有人发现！”确实，除去爱因斯坦之外，没有任何人接近广义相对论的发现，他几乎是单枪匹马地完成了这一杰作。

第二讲附录 爱因斯坦与广义相对论

1. 广义相对论的创建

爱因斯坦创建广义相对论不是偶然的，经历了长期、深刻的物理思考。在他的头脑中逐渐形成了广义相对性原理、等效原理和马赫原理等物理原理。他逐渐认识到，自己的新理论应该建立在这三条原理和光速不变原理的基础之上。

此时，爱因斯坦做出了物理思想上的一个重大突破，他大胆猜测，引力效应可能是一种几何效应，因为几何效应可以与物体的质量和组成成分无关。这样看来，万有引力可能不是一般的力，而是时空弯曲的表现。由于引力起源于质量，他进一步猜测时空弯曲起源于物质的存在和运动。

黎曼当年曾经猜测，真实的空间不一定是平的，有可能是弯曲的。现在爱因斯坦产生了与当时黎曼类似的猜想。但是，今天的爱因斯坦已经掌握了大量的物理知识，创建新理论的条件已经成熟，这些都是当年黎曼不可能具备的。

爱因斯坦1905年开始研究引力；1907年提出等效原理；1911年得到光线在引力场中弯曲的结论；1913年与格罗斯曼一起把黎曼几何引进引力研究；1915年，在与希尔伯特讨论后不久，爱因斯坦终于得到了广义相对论的核心方程——场方程的正确形式。

爱因斯坦先是与格罗斯曼合作，得到一个场方程，但有重大缺陷。方程左边表示曲率的部分与后来的正确表达式相距甚远。爱因斯坦到德国后，又与希尔伯特探讨。希尔伯特不愧是一位数学大师，爱因斯坦与他作了短时间的探讨，几个月后就给出了场方程的正确形式。希尔伯特本人也几乎同时得到了同样的场方程。

有趣的是，他们二人在最后论文的发表上曾经有过竞争。

爱因斯坦的论文是1915年11月25日投稿，当年12月5日发表的，内中包含了广义相对论场方程的正确表达式。希尔伯特投稿的时间比爱因斯坦要早，是1915年11月20日投出的，但稿中的场方程有误，在修改稿件清样期间，他看到了爱因斯坦的上述论文，于是在清样中消除了错误，该论文于1916年3月1日发表，发表时列出的场方程与爱因斯坦一致。

还有一点应该说明，希尔伯特在自己的论文投稿前一天（11月19日）曾写了一封信给爱因斯坦，祝贺他算出了水星轨道近日点进动的正确值。可见在此之前，爱因斯坦已经给出了广义相对论场方程的正确形式。

后来，希尔伯特在给爱因斯坦的一封信中称广义相对论为“我们的工作”，爱因斯坦很不高兴，回信说：“这是我的工作，什么时候成了我们的工作？”，以后希尔伯特不再提“我们的工作”，承认广义相对论是爱因斯坦的成果。此后，他们二人一直保持着真挚而深厚的友谊。

应该指出，希尔伯特只是在数学形式上得到了广义相对论场方程，并不了解它的深刻物理内容，而且，他对所得到的场方程的物理解释并不完全正确。所以完整的广义相对论理论，它的深刻物理内容和数学形式，确实主要是爱因斯坦一个人创建的。不过，希尔伯特的作用也不容忽视。从现在披露的二人当时的通信来看，那一两个月，他们二人一直在相互启发、相互促进，并最终共同走向了正确的结果。

新理论克服了旧理论的两个基本困难，用广义相对性原理代替了狭义相对性原理，并且包容了万有引力。爱因斯坦认为，新理论是原有相对论的推广，因此称其为广义相对论，而把原有的相对论称为狭义相对论。

实际上广义相对论的建立比狭义相对论要漫长得多。最初，爱因斯坦企图把万有引力纳入狭义相对论的框架，几经失败使他认识到此路不通，反复思考后他产生了等效原理的思想。爱因斯坦曾回忆这一思想产

生的关键时刻：“有一天，突破口突然找到了。当时我正坐在伯尔尼专利局办公室里，脑子忽然闪现了一个念头，如果一个人正在自由下落，他绝不会感到自己有重量。我吃了一惊，这个简单的理想实验给我的印象太深了。它把我引向了引力理论……”从1907年发表有关等效原理的论文开始，除在数学上曾得到格罗斯曼和希尔伯特的有限帮助之外，爱因斯坦几乎单枪匹马奋斗了9年，才把广义相对论的框架大体建立起来。1905年发表狭义相对论时，有关的条件已经成熟，洛伦兹、庞加莱等一些人，都已接近狭义相对论的发现。而1915年发表广义相对论时，爱因斯坦则远远超前于那个时代所有的科学家，除他之外，没有任何人接近广义相对论的发现。所以爱因斯坦自豪地说：“如果我不发现狭义相对论，5年以内肯定会有人发现它。如果我不发现广义相对论，50年内也不会有人发现它。”

2. 爱因斯坦论取得成就的原因，学校教育与“奥林匹亚科学院”

爱因斯坦在取得众多成就之后，曾经说：

“我没有什么别的才能，只不过喜欢刨根问底地追究问题罢了。”

“时间、空间是什么，别人在很小的时候就搞清楚了，我智力发展迟缓，长大了还没有搞清楚，于是一直琢磨这个问题，结果也就比别人钻研得更深一些。”

爱因斯坦不认为自己是天才。根究其做出重大成就的原因，有以下几点特别值得注意：第一，他非常勤奋，而且能够长时间地集中注意力于学习和思考。“能长时间集中注意力”这一点，不大为人注意，但却是一般人很难做到的。

第二是爱因斯坦对“奥林匹亚科学院”的高度评价。他曾经对探访他的记者说：你们为什么老问我童年和少年时代受到过什么影响？为什么不问问“奥林匹亚科学院”对我的影响？看来，爱因斯坦认为“奥林匹亚

科学院”这个自发组织的、以读书讨论为主的科学俱乐部，对自己成长为最伟大的科学家产生过重要作用。

爱因斯坦在专利局工作期间，与他的几位热爱科学与哲学的好友（先后有索洛文、哈比希特、沙旺和贝索等）组织了一个叫做“奥林匹亚科学院”的小组。这是一个自由读书与自由探讨的俱乐部。小组的成员都具有大学文化水平，他们工作单位不同，专业背景也不同，有学物理的，有学哲学的，还有学工程技术的。这几个年轻人利用休息日或下班时间，一边阅读一边讨论，内容海阔天空，以哲学为主（特别是与物理有关的哲学），也包括物理、数学和文学。他们充满热情地阅读、讨论了许多书籍，其中包括马赫的《力学史评》，这本对牛顿绝对时空观展开猛烈批判的书，对爱因斯坦建立狭义和广义相对论都产生了极大的影响。还有庞加莱的名著《科学与假设》，这本书使他们一连几个星期兴奋不已。该书内容丰富，思维活跃，其中关于“同时性”的定义、时间测量和黎曼几何的描述对爱因斯坦建立相对论可能发挥了重要影响。

爱因斯坦高度评价这个读书俱乐部，认为这个俱乐部培养了他的创造性思维，促成了他在学术上的成就。爱因斯坦曾经提醒一些记者，不要过分渲染他的童年和少年时代，希望他们注意“奥林匹亚科学院”对他的影响。

还有一个值得注意之点是爱因斯坦对学校教育评价不高。他认为学校教学方式呆板，对学生管理过严，教师居高临下地对待学生的态度，无助于学生独立精神和创造精神的培养，还会扼杀学生的自信心和学习兴趣。他觉得自己的自由创造精神未被学校教育扼杀掉，实在是个幸运。

可以说爱因斯坦一生对学校教育都没有好印象，只有对阿劳中学的看法是个例外。他回忆道：“这所学校用它的自由精神和那些毫不依赖外部权威的教师的淳朴热情，培养了我的独立精神和创造精神，正是阿劳中学成为孕育相对论的土壤。”

第三讲 白矮星、中子星与黑洞

这次讲座想介绍三种天体，白矮星、中子星和黑洞。希望大家能够从天文学的角度来了解黑洞的特点和它存在的可能性。

现在白矮星、中子星都已经看到了。白矮星是先在观测中发现，然后分析它的结构，再提出理论。中子星则是首先理论预言，后来在天文观测中发现。黑洞现在只是理论预言，还没有确凿公认地发现。虽然现在发现了很多大家觉得很可能是黑洞的天体，但是都还不确定。我记得我刚开始在北师大读研究生，七八十年代的时候，当时我们这批搞物理的人对黑洞特感兴趣，搞天文的人大都在那里观望，觉得有这东西吗？有怀疑。现在的情况倒过来了，搞天文的在那儿说到处都是黑洞，搞物理的反而有点保守，觉得这些东西真是吗？有点怀疑。

引子

我想从自己对黑洞的最早了解谈起。我第一次看到有关黑洞的叙述（当时称为暗星），知道黑洞这个概念，是在陶宏先生写的《每月之星》这本科普书中。这是“开明青年丛书”中的一本，是刚解放的时候出版的，陶宏先生在序言中说：“1949年1月22日北平停战之日写于北大红楼。”

这本书里面，陶宏先生把中国和西方的天文学对比着讲，例如一颗星，它在西方的名字是什么，在中国的名字是什么，它有些什么科学方面的内容，还做些什么科学故事和民间传说。他讲得非常好，可读性很强，其中的知识很先进。里面明确地描写了白矮星，谈到了对中子星的预言，还提到了广义相对论对黑洞的预言。《每月之星》这本书，是迄今为止我看到的写得最好的一本天文科普书，无论从科学性、知识性和趣味性，都堪称楷模。

后来我到北师大读研究生的时候，天文系主任冯克嘉先生告诉我

说，陶宏先生是陶行知先生的儿子。再后来我在师大研究生院工作的时候，跟顾明远先生在一个办公室，中午休息的时候我翻阅他的藏书，有一套《陶行知文集》，一看原来《每月之星》上的内容都是陶行知先生书上的。当时我非常感叹，这位教育家上知天文，下知地理啊，对社会也有所了解，这人真是了不起。我觉得一位伟大的教育家，应该知识面非常丰富，像孔夫子那样，基本上是一本百科全书。不是说只知道一点教育学理论就可以当教育家的。陶先生非常了不起。

《每月之星》这本书，我早年是从中学图书馆借的，中学毕业前还给学校了。前些年我又借到了一个复印本，看到陶宏先生在序言中确实讲了这是他给父亲当小助教的时候积累的资料。父亲去世了，他把它整理出来出版，大概是这样的。我随便跟大家谈一下，因为在座的很多同学都是师大的学生，希望大家知道怎样去当一名优秀的教师，怎样去争取成为教育家。我要告诉大家：“陶行知先生，上知天文，下知地理，中晓人和，全心致力于平民教育事业，不愧为伟大的教育家。”

1. 对黑洞的最早预言

拉普拉斯的暗星

我们现在来讲一下黑洞。黑洞这种东西刚开始叫“暗星”。在200多年前，拿破仑那个时代，法国的天体物理学家拉普拉斯和英国剑桥大学的学监米歇尔，几乎同时预言了这种暗星。拉普拉斯在他的书中写道：“天空存在着黑暗的天体，像恒星那样大，或许像恒星那样多。一个具有与地球同样密度，而直径为太阳250倍的明亮星体，它发射的光将被它自身的引力拉住，而不能被我们接收。正是由于这个道理，宇宙中最明亮的天体很可能却是看不见的。”他用万有引力定律进行了预言，算出了一颗恒星形成暗星的条件是

$$r \leq \frac{2GM}{c^2} \quad (3.1)$$

大家看，用牛顿的理论来看，

$$\frac{GMm}{r} \geq \frac{1}{2}mc^2 \quad (3.2)$$

按照牛顿的微粒说，上式右边是一个光子的动能， m 是光子的质量， c 是光速。左边是光子的势能，式中 M 是恒星的质量， r 是恒星的半径， G 是万有引力常数。如果光子的动能能够克服势能，远方的人就能看到光子，一般的恒星都是这样的，所以远方的人都能够看见。但是如果像式（3.2）这样，光子的动能小于、等于势能，我们就看不到恒星发出的光了。从式（3.2）不难得出拉普拉斯的结论——公式（3.1）。如果式（3.1）取等号，就可以得出“暗星”的半径：

$$r = \frac{2GM}{c^2} \quad (3.3)$$

对于太阳来说，形成暗星后，其半径是3公里，太阳现在的半径是70万公里，太阳的所有物质全都缩在这3公里范围以内，就会形成这种暗星。地球如果形成暗星，只有乒乓球那么大，整个地球的质量都缩到乒乓球那么大就成为暗星了。今天来看，上面的讨论有几个问题，一个是用了万有引力定律，没有用广义相对论；另一个问题是，光子的动能被误认为是 $\frac{1}{2}mc^2$ ，而不是 mc^2 。当然了，拉普拉斯当年不是用这个方法推导的，他用的是牛顿理论，论证的方式本质上与上面相同，但存在两个错误。非常有意思的是他却算出了正确结果。而现在用广义相对论算出来的暗星半径也是公式（3.3）。这就是说，拉普拉斯使用牛顿理论的两个错误的作用是相互抵消的，最后得到了一个正确的结果。

拉普拉斯在他的巨著《天体力学》的第一版和第二版都谈到暗星，第三版，他把暗星这一段悄悄地给撤掉了。为什么呢？因为在出版第三版和第二版之间，1801年，英国的托马斯·杨完成了双缝干涉实验，表明光不是微粒，而是波，而且是横波。这下呢，拉普拉斯用微粒说来解

释的东西，自己又觉得把握不定了，于是他悄悄把这部分内容给撤掉了。但是历史上大家都知道他曾经预言过暗星。

神童托马斯·杨

托马斯·杨是个神童，2岁就能读书，4岁把《圣经》通读了两遍。到14岁就学会了拉丁语、希腊语、法语、希伯来语、意大利语、阿拉伯语、波斯语等等，会多国的语言。他先是学医，研究近视眼，弄清了散光的原因。然后又对光学感兴趣，完成了双缝干涉实验，证明了光是波动，而且是横波，还提出了颜色的三色理论。他在十来个领域都有贡献。特别滑稽的是，他对考古学也有贡献，他把古埃及的罗塞塔石碑上的文字破译了几个，古埃及文研究的第一次突破就是他首先认出了几个字，当然没有全部突破，但是也是一个很重要的进展。他是神童，神童是个好事，但是，不是神童也没关系，牛顿和爱因斯坦都不是神童。可见，能不能有成就，最终不取决于是不是神童。所以，我觉得我们当老师的人，对学生要一视同仁，你既要鼓励那些优秀的学生，也要鼓励那些看起来好像很一般的同学。

奥本海默的预言

奥本海默是位原子弹设计师，他在搞原子弹之前，在研究中子星结构的时候（1939年）再次用广义相对论预言了暗星，也得到了与拉普拉斯相同的暗星条件。他这个发现没有引起太大的注意，没有引起大家重视，因为许多人觉得他可能在胡说。当时知道的最密的物质就是白矮星，大概每立方厘米1吨左右，在天文上观测到了这种星。这个暗星呢，太阳质量的暗星，密度为每立方厘米100亿吨，这玩意儿简直是一个让人觉得不可思议的东西！所以也就没有引起大家的注意，而且爱因斯坦也不同意会形成暗星。奥本海默后来还跟爱因斯坦在一个单位工作过。总之，这件事情当时没有引起大家更多的注意，只知道他预言过这种东西。当时都是叫“暗星”，不是叫“黑洞”。但是后来的研究表明，黑洞其实不一定密度很大，为什么呢？一个星球的密度是质量除以体积，

体积是跟半径的立方成正比的，而黑洞的半径是跟质量成正比的，从暗星半径的公式（3.3）就可以看出来，

$$\rho \sim \frac{M}{V} \sim \frac{M}{r^3} \sim \frac{1}{M^2} \quad (3.4)$$

所以黑洞的密度是跟质量的平方成反比的，质量越大的黑洞，密度越小，太阳形成黑洞，半径从70万公里缩到3公里，那当然密度很大。但是呢，假如说有 10^8 个太阳，一亿个太阳质量的黑洞的话，那密度就跟水差不多了。一会儿我们就会看到，其实谈论黑洞的密度没有什么意义。

2. 白矮星：从发现到理解

赫罗图告诉我们什么？

那么黑洞有没有可能形成呢？现在我们来看图3-1，它叫赫罗图。大家知道天上的恒星有不同的光度，这个光度指的不是我们肉眼看到的亮度，因为肉眼看到的“视亮度”取决于两个因素，一个是恒星本身的亮度，另一个是它离我们的远近。天文学家有办法测量恒星离我们的距离，他们把所有的恒星都折算到一个标准距离，这时地球上的人“看到”的亮度，称为“绝对亮度”，或者光度，它反映恒星的真实亮度、真实发光度。赫罗图的纵坐标表示恒星的光度，横坐标表示恒星的温度。有人会问怎么知道恒星的温度呢？一颗恒星我们都没去过，怎么知道温度呢？看颜色，光的颜色。一般来说，低温的恒星就发红，高温的就发蓝、发白，于是可以看出来恒星的温度。

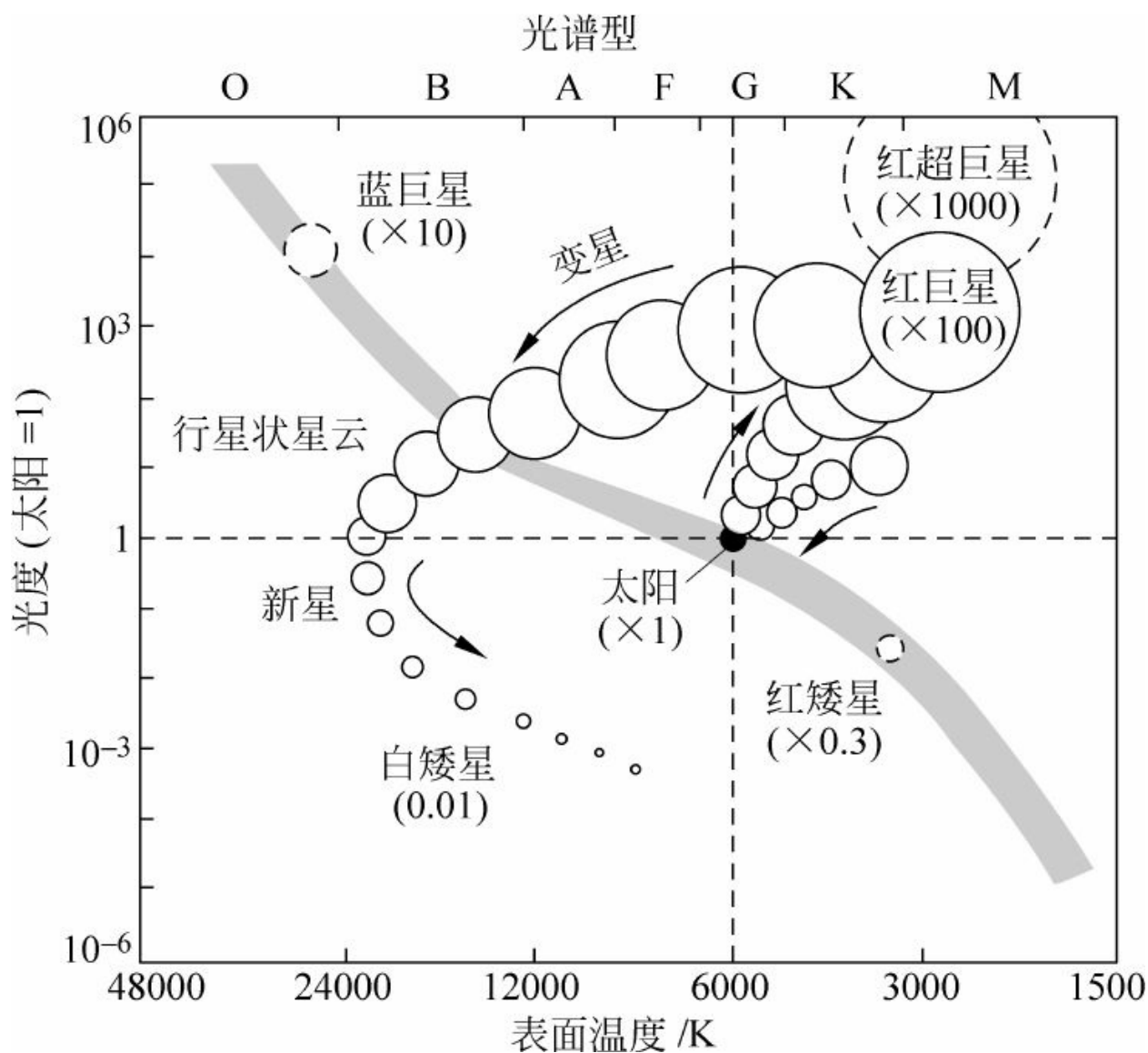


图3-1 赫罗图

纵坐标用光度，横坐标用温度，把恒星标在赫罗图上。标出来以后，就发现大多数恒星都集中在一条从左上角到右下角的带上面，这个“带”叫做“主星序”。位于主星序上的恒星称为“主序星”。我们的太阳就是一颗主序星，位于赫罗图的主星序上。还有一些恒星在主星序之外。当时对于不同温度的星，也就是不同颜色的星，根据某些特征光谱线命名了若干光谱型。比如O型、B型、A型、F型、G型、K型、M型等。

刚开始学的人觉得这东西简直太难记了。于是，有人编了一个故事。说有一个小伙子，第一次到天文台，他用望远镜一看，这么漂亮的五颜六色的星啊，哎呀，天空简直是太美了，于是就惊呼了一句：“Oh, be a fine girl, kiss me!”就是：“真像一个仙女，吻我一下吧！”这句话的每一个单词的头一个字母就代表一个光谱型，O、B、A、F、G，你把这句话记住了，光谱型的顺序就记住了。

后来的研究表明，恒星在赫罗图上的位置表示着恒星不同的年龄。主星序上的星是比较年轻的，比较老年的就会离开主星序，先变成红巨星，然后变成白矮星、中子星或黑洞。

红巨星与白矮星

现在我们就来看看主序星如何变成白矮星。像太阳这样的主序星，内部不断地进行着氢核聚合成氦核的热核反应。这种反应要在极高的温度、压强下才能进行。太阳表面温度只有6000度，但其中心有1500百万度以上的高温，还有极高的压强，在那里热核反应得以进行，不断地放出能量。太阳这类恒星（主序星）会维持这种状态相当长一段时间，在相当长一段时间内就这样发光，比较稳定。但是，中心部分的氢总会烧完，烧完以后就开始烧外层的氢，这时候这颗恒星就慢慢膨胀起来，形成红巨星。太阳形成红巨星时会扩展到把水星、金星都吞到它肚子里头，把地球上的江河湖海都烤干，然后也吞进去，一直伸展到火星的轨道，形成那么大的一颗红巨星。红巨星温度比较低，大概是4000度，所以发红。

大家想，唉吆，那我们不就完了吗？是不是？世界末日不就降临了吗？不过大家可以放心，根据现在的研究，太阳在目前这种状态的寿命应该能维持100亿年，现在已经过了多少？过了50亿年。所以大家尽可以放心地活着，还有50亿年基本上是现在这种状态。50亿年之后，人类的科学会高度发达。大家想一想，从哥白尼到现在，现代自然科学才500年，我们已经可以登月了，你想50亿年之后的人类会怎么样？所以

大家对未来可以充满信心。

这种红巨星的中心部分会缩成白矮星。基本上就是这样，主序星，就是主星序里面的这些恒星，都是我们太阳现在的这种状态，然后它们将演化成红巨星，然后变成白矮星，再冷却成黑矮星。黑矮星就是一块巨大的金刚石，主要由碳和少量氧构成，在天空中飘荡，我们要是弄到一块，那就发大财了，是吧？可是到现在一颗黑矮星都没有找着，因为白矮星冷却到黑矮星要100亿年，宇宙今天的年龄才130多亿年，作为大金刚石的黑矮星，大概一颗都还没有形成，还需耐心等待。

大一点的恒星，例如超过太阳质量10倍左右的恒星，会演化成超红巨星，然后会超新星爆发，最后形成中子星、黑洞或者全部炸光。

图3-2是恒星的演化图。

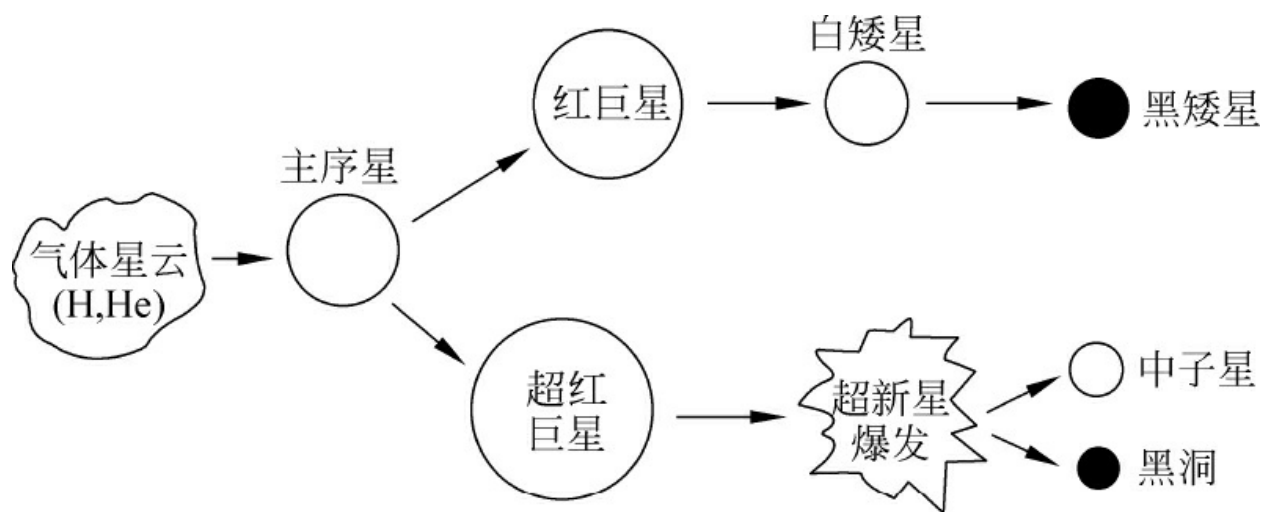


图3-2 恒星的演化

各种恒星与黑洞的比较

我们现在来比较一下。太阳现在的半径是70万公里，密度是每立方厘米1.4克，跟水差不多。它将来形成白矮星时半径1万公里，密度是每立方厘米1吨。它如果形成中子星，半径10公里，密度是每立方厘米1亿到10亿吨。如果形成黑洞，半径3公里，每立方厘米100亿吨。我们太阳最后的结局是白矮星，不会形成中子星和黑洞，因为它质量不够大，不

会超新星爆发。

图3-3是一个示意图，它大致告诉我们，太阳形成红巨星和白矮星后会有多大，如果形成中子星和黑洞又会有多大。太阳、红巨星、白矮星和中子星之间体积的差异都很大，中子星和黑洞体积的差异则较小。现在红巨星、白矮星、中子星都已经发现了，大小和中子星相差不大的黑洞，似乎没有理由不存在。

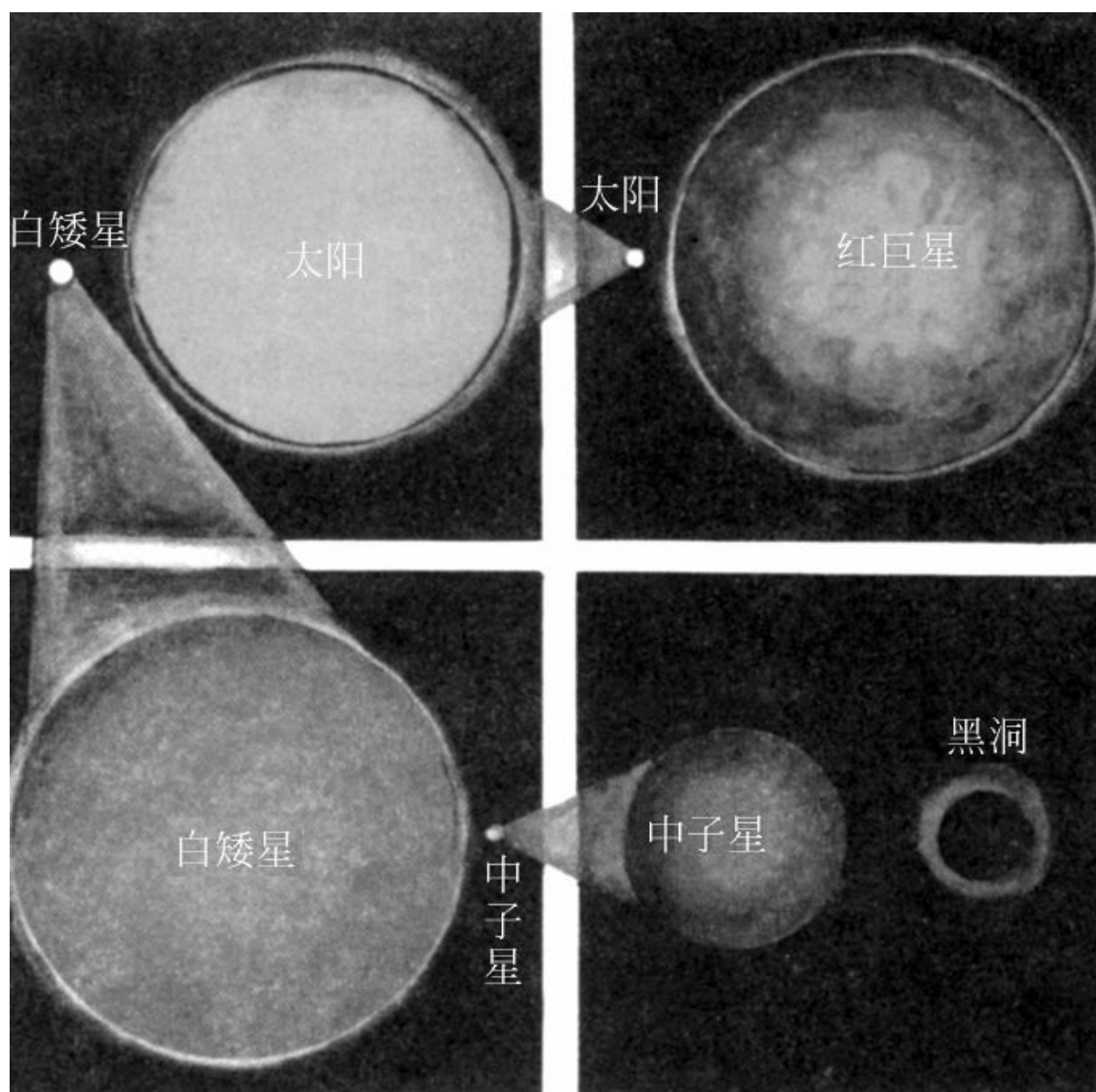


图3-3 各种恒星大小的比较

狐矢射天狼（图3-4）

最早看见的白矮星是天狼星的伴星。大家看，图3-4这颗星是大犬座的 α 星，这是国外的名字。中国人管它叫天狼星，西洋人叫狗，我们叫狼，差不多。中国古代认为天狼星代表侵略。左下方这一组星叫做弧矢星，弧矢就是弓箭，弧矢射天狼，就是反击侵略者。

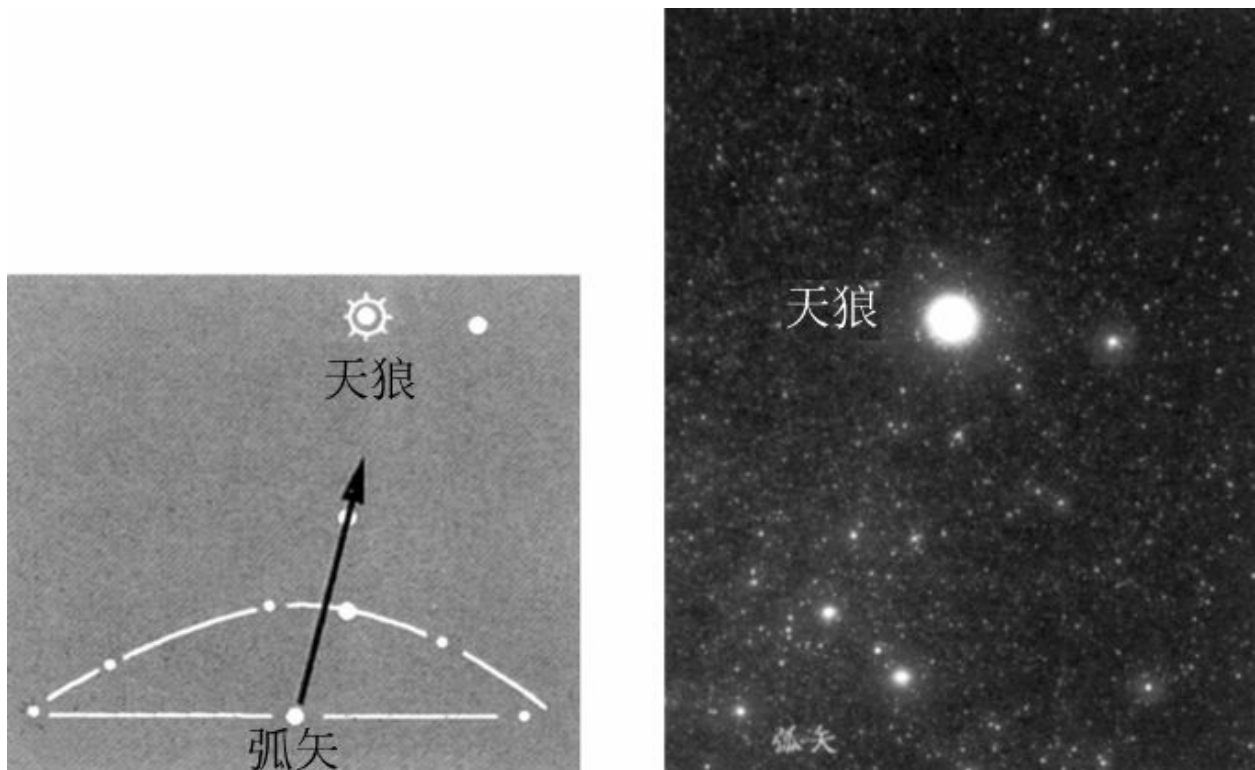


图3-4 弧矢射天狼

所以屈原的诗里就有：“举长矢兮射天狼，操余弧兮反沦降。”抗日战争的时候，大后方的文人经常引用这句诗。北宋的苏东坡也说：“会挽雕弓如满月，西北望，射天狼。”为什么“西北望，射天狼”？一个是在天空中，天狼星在弧矢星的西北。另外一个就是北宋的主要敌人是西夏。北宋跟辽国虽然在初期有战争，但是檀渊之盟订立了一个对宋朝来说比较屈辱的和约之后，北部的宋辽边境安宁了下来。此后西北方的西夏取代辽国，成为北宋的主要敌人。

天狼星的神秘伴星

我们来解释一下恒星为什么叫恒星。这是因为，从地球上看来，它们在天空的相对位置都不变，形成固定的星座。不像太阳、月亮以及金木水火土五颗行星，它们在恒星形成的天空背景上移动，穿过各星座移动。天上的恒星为什么都不动？其实它们都是银河系里边的星，都在围着银河系的中心转，只不过离我们远，所以我们就觉得它们不动。如果能把人冰冻起来，10万年以后再让他缓过来，去看天上的星，就不一样了。假如他现在认识天上的星座，那时候就不认得了。恒星在天空的相对位置都变了。这就是说，恒星并非真的不动，只是它们离我们太远，看起来它们动得很慢。

不过天狼星离我们比较近，有多近呢？9光年，就是说光走9年就到了，很近。所以天文学家早就发觉它在天空绕一个小圈。于是，有人推测它是不是有一颗伴星，否则为什么我们看见天狼星老在那绕圈呢？肯定是有个东西吸引着它，二者围绕它们的质心在转。有人打个比方说：“你看，现在有一个小伙子跟一个姑娘在跳舞，小伙子穿一身黑色的礼服，姑娘穿一身白色的连衣裙。当光线逐渐暗下来以后，你就看不见那个小伙子了，只看见那个穿白衣服的姑娘在那儿转。她围着什么转呢？有个东西吸引她。”后来发现这个吸引天狼星的星虽然个子小，但是密度很大，温度很高，颜色发白，因此称它为白矮星（图3-5）。这是最早发现的一颗白矮星。如今我们已知道，宇宙中的白矮星很多，大约占全部恒星的十分之一左右。

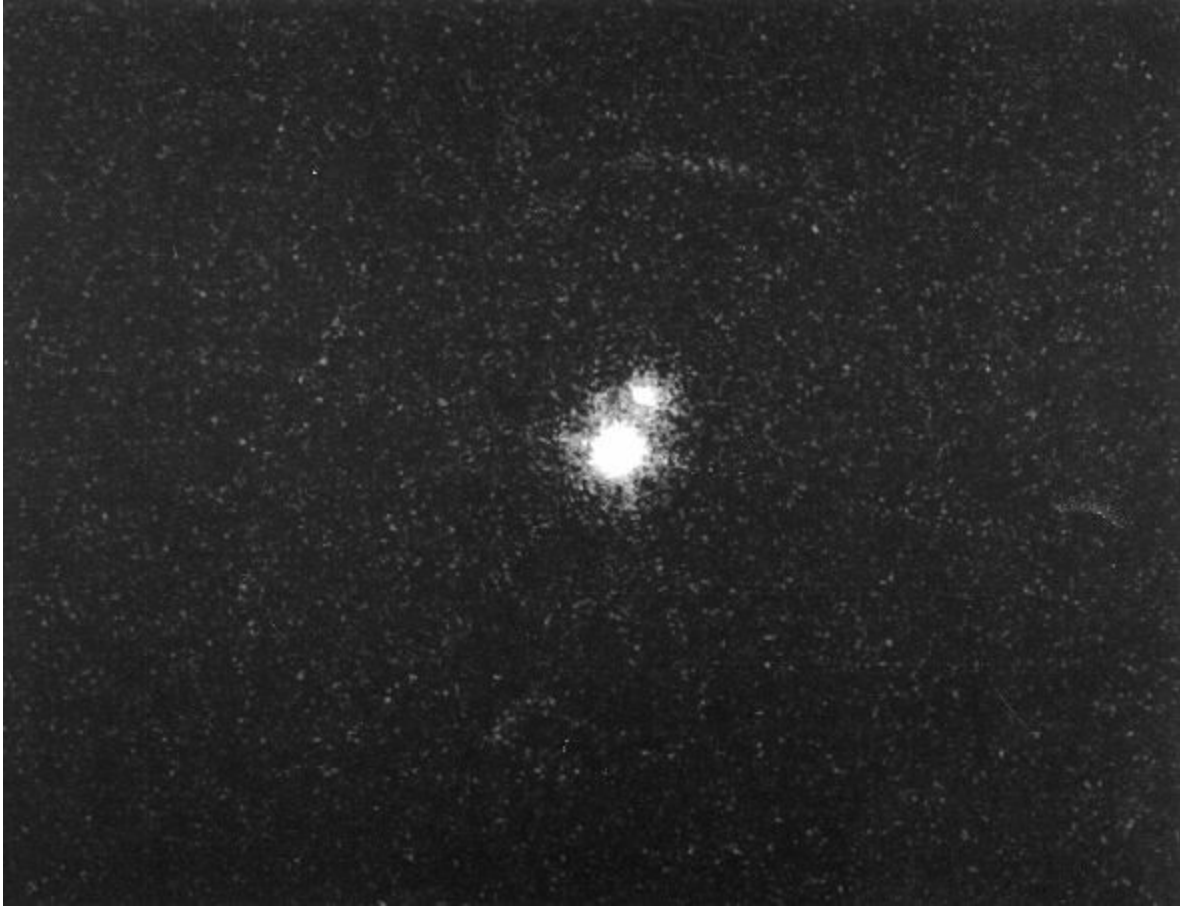


图3-5 天狼星和它的伴星，伴星是白矮星

霍伊尔的功绩

白矮星是怎样形成的呢？我们知道，恒星演化的主序星阶段，是氢通过热核反应燃烧生成氦，中心部分的氢烧完就烧外层的氢，于是外层膨胀，形成红巨星，内层往里面收缩。在收缩过程中，星体中心部分的温度迅速升高，那里有大量的氦和少量的氢，它们能进一步聚合吗？研究表明，两个氦核聚合的生成物（共8个质子与中子），或者一个氦核与一个氢核聚合的生成物（共5个质子与中子）均不稳定，这样的聚变反应不可能发生。3个氦核聚合在一起生产的碳（6个质子与6个中子）倒是稳定的，但3个核同时碰在一起的概率很低，这样的反应似乎更难发生。

天体物理学家霍伊尔猜测，碳核可能存在一种激发态，其能量恰好

与3个氦核加起来的总能量相等，这时在3个氦核与激发态碳核之间会发生一种“共振反应”，使聚合概率大大提高。生成的激发态碳核会很快跃迁到基态，形成稳定的碳元素。这样氦聚合成碳的热核反应就得以进行了。一些核物理学家最初不相信霍伊尔的猜测，但他们查找后，真的发现了这种碳核的激发态，确认了“共振反应”的存在。大家终于明白了，通过“共振反应”，氦将进一步聚合生成碳，并释放出大量核能。而且碳还可以与氦再进一步聚合成氧。

白矮星主要由碳和少量氧组成，它的密度大约为每立方厘米1吨左右，它怎么才能维持住自身不往下塌呢？我们知道，固体的行星不往下塌，靠的是电磁力支撑。万有引力使构成固体的原子相互靠近，电荷分布发生变化，同种电荷间的斥力支撑着它不往下塌。白矮星密度这么大，电磁力抗拒不住万有引力，会把原子的电子壳层压碎，形成原子核的框架在电子的海洋当中漂浮的状态，或者说电子在原子核形成的晶格中穿行的状态。这时电子靠得很近，泡利不相容原理的排斥力开始起作用，支撑住它不往下塌。这就是白矮星的物质状态。

钱德拉塞卡的发现

因为时间关系我就不讲那么多了。我现在讲另外一件事情，讲一下印度的一位著名的科学家钱德拉塞卡的贡献。他首先指出，白矮星还有可能再往下塌。他认为，白矮星有一个质量上限，1.4个太阳质量。超过1.4个太阳质量的白矮星肯定不稳定，电子间的泡利斥力顶不住万有引力，要继续往下塌。他当时二十几岁，在印度大学毕业。

印度以前高等教育比我们要发达，但是这些年中国的高等教育发展大大地加速了，这是印度人没有想到的。前几年温总理在德里理工学院（印度最好的理工科大学）演讲的时候，有一个印度学生问他：“中国的高等教育什么时候能够赶上印度的水平？”温总理没有正面回答他，就告诉他中国有多少大学生，这么一说，全场的印度学生都惊呼起来了。想不到中国有这么多的大学生。

20世纪20年代，印度已经有了比较好的大学。钱德拉塞卡在印度大学毕业，他喜欢天文，想到英国学习天体物理。他坐船去英国，在海上漂了约20天。他每天躲在船舱里计算。到达英国的时候他算出了新结果：白矮星有个质量上限，超过这个质量上限，泡利斥力就顶不住了。他与一些天文学家进行了讨论，确认无误后，又去请教著名天体物理学家爱丁顿，没想到爱丁顿说他的计算肯定不对。爱丁顿想，物质再塌下去不就缩成一个点了吗？这怎么可能呢！钱德拉塞卡很有绅士风度，爱丁顿不同意，他就等了一段时间，又来找爱丁顿，说：“爱丁顿教授，我没有看出我的计算有什么错，您是不是再看看？”爱丁顿说：“你肯定错了！”就这样反复了几次，爱丁顿想这家伙怎么回事啊，老跟我说这事儿！就对钱德拉塞卡说：“过两个星期，在伦敦有一次学术讨论会，我给你争取一个发言机会，还可以给你争取双倍的发言时间。”因为学术讨论会上的发言都是有时间限制的，比如说，一个人可以讲20分钟。有人说我要讲两个钟头，那不行。你有时间讲，别人还没有时间听呢。因此每个人的发言时间是很有限制的。现在，爱丁顿为了让钱德拉塞卡有足够的时间讲清楚自己的理论，所以说要给他争取双倍的发言时间。

爱丁顿的反对

开会的前一天晚上，钱德拉塞卡与爱丁顿一块吃饭，他问爱丁顿：“爱丁顿教授，您明天也有报告吗？”爱丁顿说：“有。”“报告的题目是什么呀？”“跟你的一样。”当时钱德拉塞卡就紧张了。他想：“爱丁顿是不是要篡夺我的研究成果啊！”第二天他报告之前给大家发了预印本，所谓预印本就是没有正式发表的论文，先印出来给别人看的。他讲完之后，爱丁顿拿着他的一份预印本上去了，说：“刚才钱德拉塞卡那个报告我认为全是胡扯，完全是错误的。”“噌噌”就给撕了。哎呦，当时弄得钱德拉塞卡非常难堪。钱德拉塞卡的一些朋友在散会以后跟他说：“糟透了，简直糟透了！钱德拉塞卡，这次糟透了！”大家觉得钱德拉塞卡闹了个大笑话。爱丁顿为什么那么气粗呢？因为他把钱德拉塞卡

的观点告诉过爱因斯坦，爱因斯坦也同意爱丁顿的意见，可见伟人也不见得不犯错误。

泡利的支持

钱德拉塞卡很狼狈，后来去了美国，在美国找了个工作。他24岁时提出这个理论，73岁的时候因为这个发现获得了诺贝尔物理学奖。不过在此之前，他已经知道自己的理论是对的。为什么呢？有一些人表示赞同。特别是有一次他去向泡利请教。

泡利是聪明得不得了的一个，他19岁就写了一本广义相对论讲义，22岁的时候这本广义相对论讲义就正式出版了，书名叫《相对论》，我还有他这本书。就是今天来看，这本书也是高水平的。泡利真是不得了，他总觉得别人都不行，就他行，很自负。别人在那里做报告，他就在那儿挑毛病，反正一般人都被他挑得害怕。有位年轻人在研究一个问题，他的朋友告诉他：“你知道不知道，泡利最近也对这个问题感兴趣，这可不是一个好消息。”杨振宁先生年轻时就曾经几次被泡利追问问题。

有一次开会，钱德拉塞卡见泡利也在场，就跑去问泡利，说：“泡利教授，您觉得我这篇论文怎么样？”泡利说：“很好啊！”泡利很少赞扬人，这次却赞扬说：很好啊。钱德拉塞卡说：“爱丁顿教授说我的结论不符合您的不相容原理啊。”他说：“不不不，你这个结论符合泡利不相容原理，可能不符合爱丁顿不相容原理。”他把爱丁顿讽刺了一通。

3. 中子星：从预言到发现

那么，超过钱德拉塞卡极限的星体塌下去会怎样呢？会像爱丁顿和爱因斯坦他们顾虑的那样塌成一个点吗？不会。研究表明，这时候电子会被压进核里面去，跟原子核里的质子“中和”，形成大量中子，成为一个以中子为主体的恒星，叫做中子星。

发现中子的曲折

中子星是首先预言，后来才发现的。1930年普朗克的研究生玻特发现了一种看不见的、不带电的、穿透力很强的射线。1931年约里奥夫妇，就是居里夫人的大女儿和大女婿，研究了这种射线，他们都认为是 γ 射线。约里奥夫妇化学水平很高，但是他们的物理水平不是很高，因为他们俩都是学化学出身的，他们头脑中没有中子这个概念。

这个时候，英国有一人正在找中子，就是卢瑟福的学生查德威克。查德威克的老师卢瑟福早就注意到许多原子的原子量和原子序的差近乎整数，于是他一直猜测，是不是原子核里有一种未知的粒子，质量跟质子一样，但是不带电，也就是今天所说的中子。查德威克一直在找中子，这时候看见了约里奥的论文，高兴得不得了。“哎呦，他们看见了中子还不知道啊！”于是他马上做了一个类似的实验，在英国的自然杂志《Nature》上刊登了，说中子可能存在。然后又写了一篇长文章在英国皇家学会会报上登出，题目是《中子的存在》。于是中子就被发现了。

这时约里奥夫妇感到非常地懊丧，到手的发现自己没有抓住，正应了法国生物学家巴斯德的话“机遇只钟情于那些有准备的头脑”。他们两个人的头脑没有准备，错过了这一发现。但是他们俩并不气馁，继续探索，不久之后用人工方法造出了放射性元素。此前，人类用的放射性元素都是天然形成的。

1935年颁发诺贝尔奖，大家认为中子的发现应该获奖。当时有人主张约里奥夫妇和查德威克分享诺贝尔奖，但是评委会主任卢瑟福是查德威克的老师，他说：“约里奥夫妇那么聪明，他们以后还会有机会的，这次的奖就给查德威克一个人吧！”于是查德威克获得了诺贝尔物理学奖。当年的下半年，同一个评委会评化学奖，就把化学奖颁给了约里奥夫妇，理由是他们发现了人工放射性，即人工造出了放射性元素。

当时物理奖跟化学奖分得不太清楚，例如卢瑟福本人也得的化学奖。得奖通知来的时候，他也没想到会给他一个化学奖。卢瑟福打开信

一看，就哈哈大笑，说：“你们看呐，他们给我的是化学奖，我这一辈子都是研究变化的，不过这次变化太大了。我从一个物理学家一下变成化学家了。”

中子星的预言

书归正传，我们接着讲有关中子星的事。1932年中子发现的时候，消息传到哥本哈根的理论物理研究所，当天晚上玻尔就召集全所的人开会，也就是二十几个人吧，让大家畅谈一下中子发现的感想。当时有一个在那进修的年轻的苏联学者，叫朗道，立刻即席发言，说：“宇宙中可能存在主要由中子构成的星。”也就是中子星，这是我们知道的最早的对中子星的谈论。朗道这个人水平高得不得了，按照杨振宁的说法，20世纪最棒的物理学家就是爱因斯坦、狄拉克和朗道。

1939年，奥本海默对中子星进行理论研究时，发现中子星也有一个质量上限，大约3个太阳质量。超过这个质量上限（即奥本海默极限）的中子星还要继续往里面塌，于是奥本海默预言了“暗星”，也就是我们今天所说的黑洞的存在。

“小绿人”

1967年，英国剑桥大学的休伊士和贝尔发现了中子星。他们是偶然发现的。休伊士设计了一套接收宇宙中来的无线电波的仪器装置。我们看见的恒星都是发可见光的，但是有一些天体是既发可见光又发无线电波的，还有一些是只发无线电波不发可见光的。他设计了一套装置，然后让他的女研究生贝尔在那里作巡天观测，寻找各种射电源，就是无线电波的发射源。有一个假日，休伊士回家了，贝尔在观测中突然发现，有一种很规则的脉冲信号。她就赶紧打电话叫她的老师过来。老师过来看了以后说：“这个确实值得注意，你不要告诉任何人。”然后他就研究了，公布说发现了一个这样的射电源，别人就问：“在哪儿？”打电话问他，他不告诉。又宣布发现了一个，然后又发现了一个，别人问他，他都不说。结果有的人就发火了：“有你这样搞研究的吗？你既然公布了

就应该告诉我们在哪儿，大家共同研究嘛！”最后在压力之下，休伊士被迫说出了几个射电源的位置。诺贝尔奖评委会那次奖发得很不正确，只给了休伊士，没有给贝尔。这件事情在天体物理界引起了轩然大波，很多人出来为贝尔打抱不平。休伊士说了一些很不应该说的话，他说：“那怎么了，奖发得没什么问题啊。这仪器是我设计的，是我让她看的。”可是话又说回来，休伊士给贝尔布置的研究任务并不是寻找中子星。实际上这是一个计划外的偶然发现。如果贝尔不仔细看，不认真，不认为这是个值得注意的信号的话，那他们就发现不了。所以有些人对休伊士很有意见，包括霍金，他们开始挖苦休伊士。特别是霍金，在他的《时间简史》书里有一张贝尔一个人的照片，强调中子星的发现者首先是贝尔，其次才是休伊士（图3-6）。

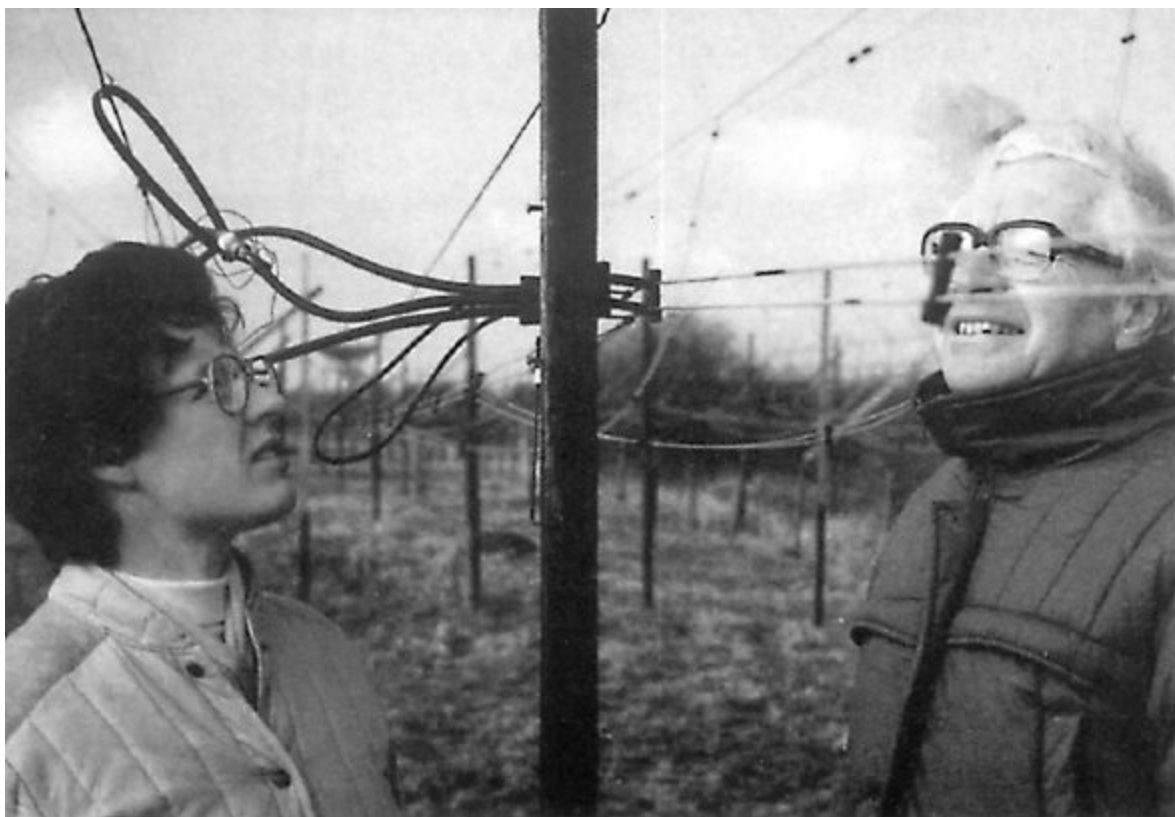


图3-6 脉冲星的发现者——休伊士和贝尔

贝尔曾几次来中国。这个人人品非常好，对她的老师从来没有一句

抱怨，只是她改行干别的了。别人怎么说这事，她都不附和，不吭气。越这样呢大家越同情她。中国曾经多次邀请她来访问。

休伊士和贝尔刚开始收到这种规则脉冲信号的时候，以为是外星人跟咱们联络呢，就取了个代号叫“小绿人”。后来发现这个“小绿人”根本不是外星人，为什么呢？这些脉冲没有任何变化，间距没有变化，振幅也没有变化，不负载任何信号。后来发现这是中子星发射的脉冲，是高密度的中子星旋转的时候射出来的。

脉冲星

你们看图3-7，这是一个高密度的旋转中子星，它磁场非常强，为什么？你想太阳有磁场，它如果体积收缩的话磁场并不会消失，在那么小的范围之内磁场会显得很强，强磁场引起很多电子旋转，就会沿着两个磁极方向产生辐射。中子星转动的速度非常快，每秒钟能转几百转。怎么会转那么快呢？因为角动量守恒。恒星是有自转的，但转速一般很慢，这一塌缩下去，体积大大减小，转动惯量减小了，但要保持角动量守恒，转速就必须增大。由于自转轴一般不与磁轴重合，沿磁轴的电磁辐射就像探照灯的光柱似的在宇宙空间扫描，每扫过地球一次我们就收到一个脉冲，再扫过一次又收到一个脉冲，所以这种星起初叫脉冲星。现在已经知道了，脉冲星其实就是中子星，密度每立方厘米大概1亿吨到10亿吨。目前中子星的发现已经被科学界确认了。

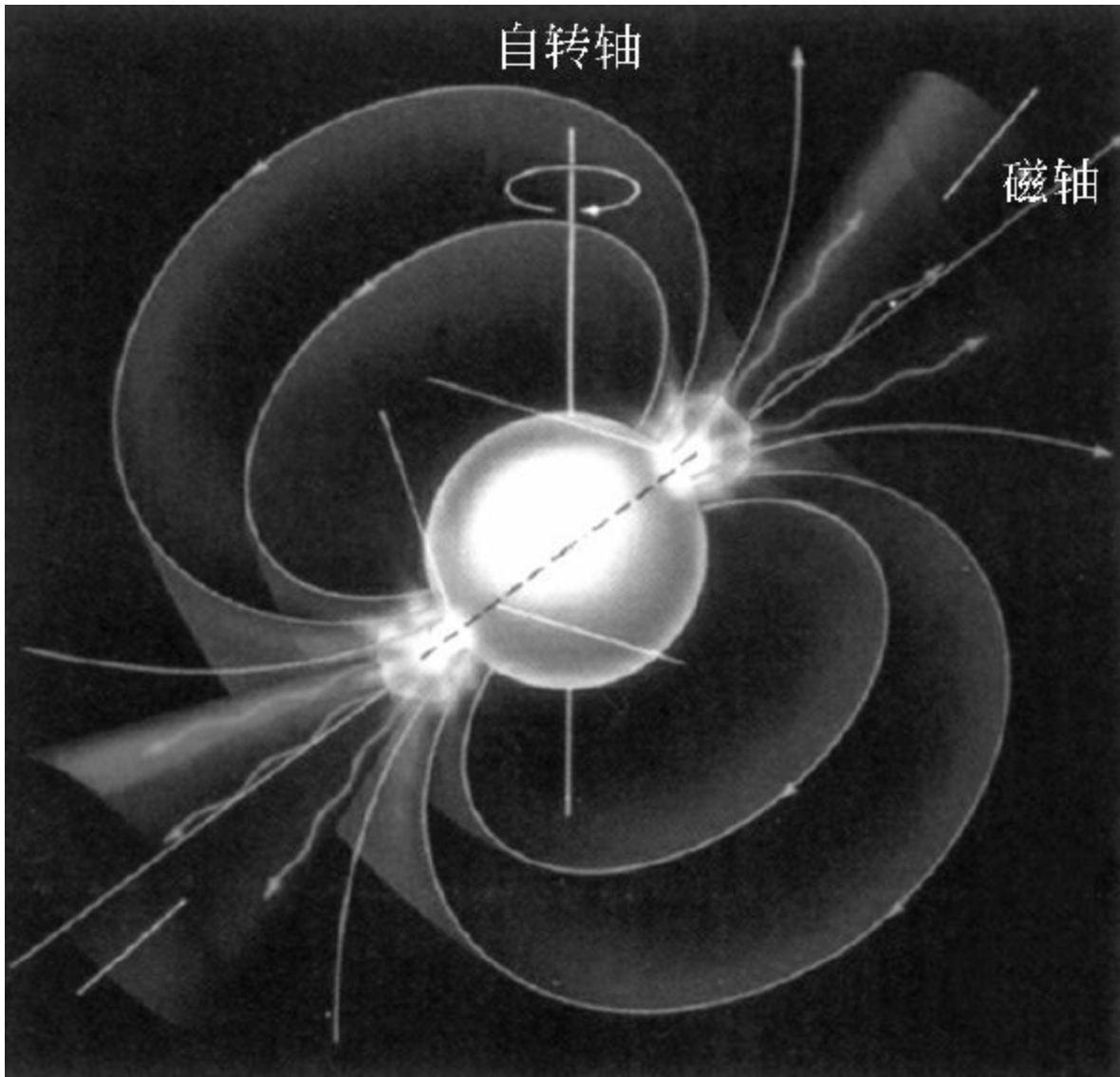


图3-7 脉冲星—中子星

我们中国人对于中子星的发现也是有贡献的，那是在宋朝的时候。宋仁宗至和元年（公元1054年），我们的祖先看到，天空中出现了一颗“客星”。所谓“客星”就是在本来没有恒星的天空位置上，突然出现的一颗亮星。有多亮呢？“昼见如太白”，即白天看起来像金星那样亮，持续了23天，然后暗下去，但此后有一年多的时间，夜间仍可看到。日本人、越南人、朝鲜人也看到了这颗“客星”，但只有中国人记录了这颗客星的方位（图3-8）。

凡十一日没三年三月乙巳出東南方大中祥符四
 年正月丁丑見南斗魁前天禧五年四月丙辰出軒轅
 前星西北大如桃速行經軒轅太星入太微垣掩右執
 法犯次將歷屏星西北凡七十五日入濁没明道元
 年六月乙巳出東北方近濁有芒彗至丁巳凡十三
 日没至和元年五月己丑出天關東南可數寸歲餘
 稍没熙寧二年六月丙辰出箕度中至七月丁卯犯
 箕乃散三年十一月丁未出天囷元祐六年十一月
 辛亥出參度中犯掩側星壬子犯九游星十二月癸
 酉入奎至七年三月辛亥乃散紹興八年五月守婁

图3-8 中国人对超新星的记载

现在我们知道，这种客星就是恒星演化到晚期发生的大爆炸现象，叫做“超新星爆发”。质量超过太阳七八倍以上的恒星晚期都会出现超新

星爆发。超新星一天发出来的光，相当于太阳一亿年发出来的光，它的亮度几乎可以与整个星系（含有大约10亿颗恒星）的亮度相比（图3-9）。

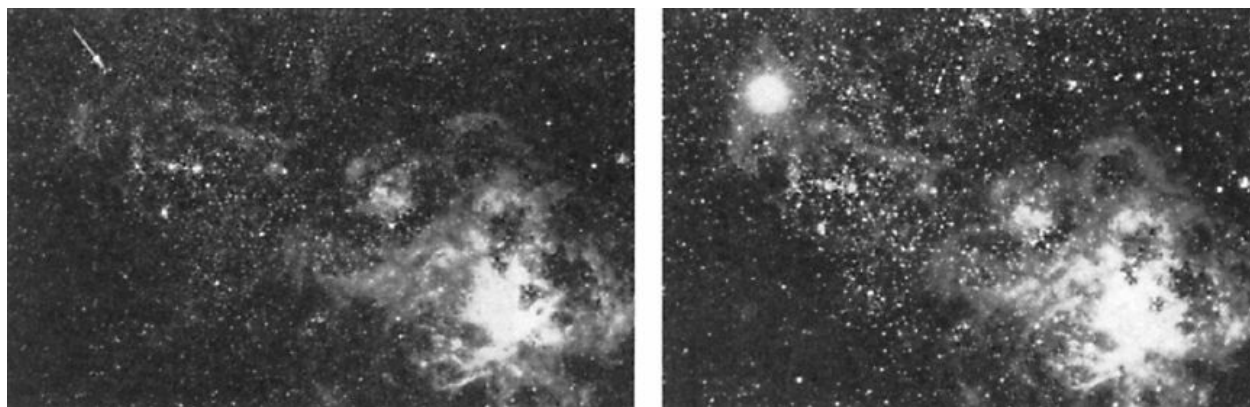


图3-9 超新星爆发

1731年，英国的一个天文爱好者（职业是医生），在金牛座发现一个螃蟹状的星云（图3-10）。1928年，哈勃认识到它由气体和尘埃构成，正以大约每秒1100公里的速度膨胀，其中心有颗小暗星，推测它们是超新星爆发的遗迹。1944年，一位天文学家和一位汉学家合作，认识到这个蟹状星云就处在中国人记录的超新星爆发的位置，这些气体和尘埃就是这颗超新星爆发的遗迹。中心的暗星就是爆炸的残骸。1968年，发现这颗暗星是一颗脉冲星，也就是中子星。这一发现证实了人们的推测，中子星是经过超新星爆发形成的，爆发的结果不仅能形成中子星，还可能形成黑洞。



图3-10 蟹状星云

中子星靠什么支撑呢？白矮星是靠电子之间的泡利斥力，中子星是电子压到原子核里后与质子结合形成中子，它是靠中子之间的泡利斥力支撑的。中子之间的泡利斥力比电子之间大，但也不是可以无限大，所以中子星也有个质量上限。这个质量上限叫奥本海默极限。超过这个极限，中子间的泡利斥力就顶不住了，中子星就进一步塌缩最后形成黑洞。

4. 黑洞初探

奇点与奇面

我们现在介绍一下黑洞。拉普拉斯和奥本海默都谈到过形成暗星的条件。暗星的半径如式（3.3）所示，物质都缩到里面以后就形成了黑洞。

图3-11中这个半径 $r = \frac{2GM}{c^2}$ 的球面后来起名叫视界，这是一个奇异的表面。在中心 $r=0$ 这个地方，还有一个奇点。

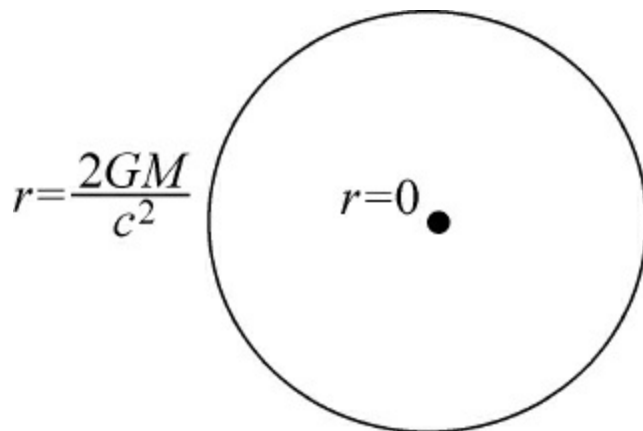


图3-11 球对称的黑洞

“时空互换”

黑洞很有意思，在它的里边，时空坐标会互换，也就是说在它里边 t 坐标不再表示时间，而是表示空间了。而它的那个半径 r 变成了时间。时间是有方向的。黑洞内部的时间方向是朝里的，朝中心的。所以进到黑洞里的物质都不能停留，要一直缩到奇点上面去。因此黑洞没有什么高密度的结构，它里面都是些真空。外边可以是真空，当然也可能有物质围着它转。但是物质一旦掉进去以后就会直奔这个奇点，只有奇点处的密度是无穷大。奇点问题的研究现在还不是特别清楚。

r 现在是时间了，那么这个 $r=0$ 就不是球心了。只有 r 是空间坐标，是半径时， $r=0$ 才是球心。现在 $r=0$ 是时间的终点，是时间结束的地方。黑洞内部时间方向朝里。有人问时间为什么朝里，它为什么不能朝外？可以朝外，朝外就是白洞。相对论只告诉了我们是有洞，并没说它是黑洞还是白洞。为什么大家一般只谈论黑洞呢？因为这种东西是高密度

的星体往里塌缩形成的。它初始条件是向里掉的，所以我们谈的都是黑洞。但是相对论并不否认白洞，也有一部分论文谈论白洞，照样可以发表。但是人们想不出来自然界怎么才能形成白洞。

飞向黑洞的火箭

现在我们就来看，当有一个火箭飞向黑洞的时候，远方的观测者能看到什么？火箭上的人能感受到什么？图3-12有一个黑洞，这个箭头表示一个火箭，远方有一个人看着这个火箭飞向黑洞，他能看见什么？

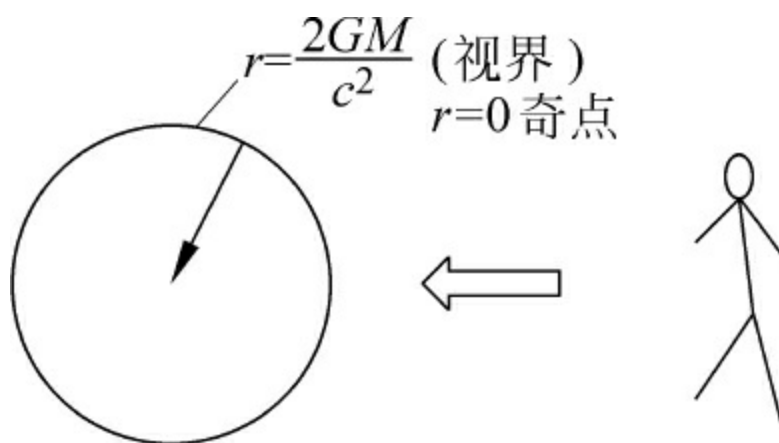


图3-12 飞向黑洞的火箭

我们知道，在物质密度特别大的地方时空弯曲得厉害，时空弯曲得厉害的地方，时钟就走得很慢。所以越靠近黑洞表面，放在那里的钟走得越慢。根据现在的研究，如果有一个钟摆在黑洞表面，那根本就不走。远方的人看它，根本就不走。所以远方的人将能看到火箭越飞越慢，越飞越慢，最后就粘在黑洞的表面上，进不去。还有什么现象呢？这火箭会越来越红。为什么呢？因为有红移，时钟变慢后就会出现红移。远方的人就看见这火箭是越来越慢，越来越红，最后就粘在黑洞的表面上。还有什么呢？会看到这火箭越来越暗，消失在那个地方的黑暗之中。但看不见它进去。

进入黑洞的冒险者

那么它进去没有呢？它进去了。火箭上的人用的钟，不是远方观测

者的那个钟，而是他自己携带的火箭上的钟。他觉得自己很顺利地就进入黑洞了。并且进去以后火箭不能停留，而会直奔奇点。这是因为黑洞里面“半径” r 成了时间，时间有方向，会不停地向奇点处流逝，所有进入黑洞的物体必须“与时俱进”，奔向奇点，不能停留，所以它的末日就降临了。奇点那儿密度是无穷大，当火箭非常接近奇点的时候，潮汐力就会把它撕碎。

既然火箭进去了，为什么外边的人没看见它进去呢？这是因为它的背影留在外头了。我们地球上有人出门，一出门你看他那背影一闪，没了。为什么没了呢？组成他背影的光子都过来了，不再存在了。可是黑洞的表面呢，那里时空弯曲得很厉害，组成火箭背影的光子不会一下都跑出来。它们会慢慢地往外跑，越跑光子密度越稀，越跑越稀，所以你看到的背影是慢慢消失的，越来越暗，越来越暗，然后看不见了，但是你看不到它进去。因此远方的人看见火箭是越来越慢，越来越红，越来越暗，最后粘在黑洞的表面上，像冰冻一样冻结在黑洞的表面，消失在那里的黑暗之中。所以苏联人管黑洞叫“冻结星”。

苏联的泽尔多维奇，那是个很了不起的人。他刚开始只是物理实验室的一个实验员，后来科学家们发现这个小伙子虽然没上过大学，但是太聪明太能干了。就鼓励他去学习，让他进修，后来他成为苏联最杰出的理论物理学家、天体物理学家，是苏联核武器的主要设计者之一。最早研究黑洞的人许多是核武器的设计者。因为原子弹制造完以后，不知道该干什么了，就正好研究黑洞。

那么，进入黑洞的人有什么感觉呢？他会感到潮汐力越来越大。什么是潮汐力呢？潮汐力就是万有引力的差。比如说我站在地面上，我受到一个重力，我脚底受到的重力和头顶受到的重力大小是有点差别的。原因是什么呢？就是我脚底到地心的距离和头顶到地心的距离差一个 δ 。这个 δ 呢，就是我的身长。对不对？这个重力差有多大呢？3滴水的重量，所以我们平常都感觉不到。

为什么地球上会有涨潮落潮呢？主要是月球的引力造成的。你看图3-13，这是一张示意图，这是地球，有一个水圈。A点到月亮的距离和B点到月亮的距离，差地球的直径，所以这两点有个万有引力差，A、B这两个方向是涨潮，其他方向是落潮。当然还有一个次要作用就是太阳的因素。太阳和月亮如果都在地球的同一侧，或者在两侧，那么就涨大潮。如果日地连线是在月地连线的垂直方向上，那么正好两个作用有点抵消，涨的就是小潮。当然，要具体研究这个问题，还是需要做一些计算的。这只是一个示意图。

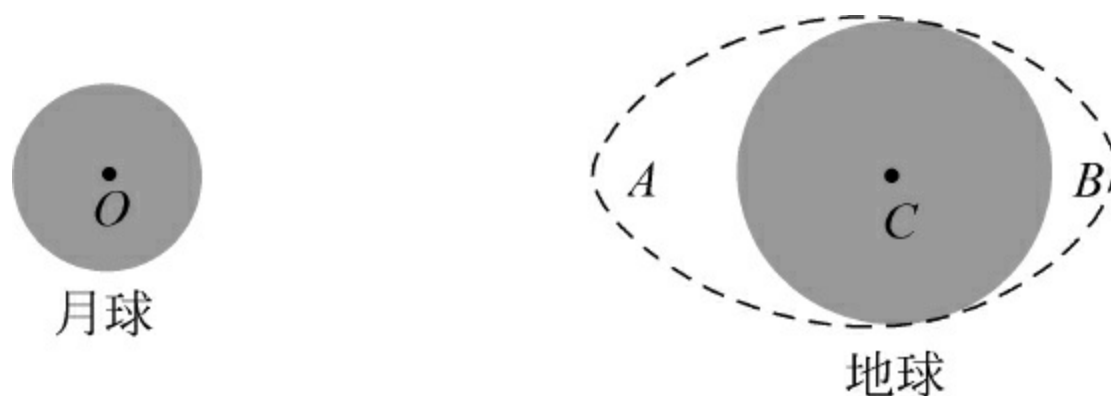


图3-13 涨潮与落潮

在黑洞里，火箭靠近奇点的时候，火箭前后受到的潮汐力非常大，会把火箭和宇航员全部撕碎，然后压到奇点里面去。奇点是时间的终点，于是火箭和宇航员就处于时间之外了。什么叫处于时间之外呢？这事还不清楚，还不能给出很好的解答，现在学物理的人也不进行探讨，觉得不可能处于时间之外，但是也不清楚这个情况到底是怎么回事。

转动的黑洞

现在，我们来看一下转动的黑洞（图3-14）。一个旋转的黑洞，里面的结构是很复杂的。对于不旋转的黑洞，飞船进去以后肯定就是它的末日了。但是旋转的黑洞呢，它里面有个空间，火箭还可以在里头转，中间不是一个奇点，而是一个奇环。因为黑洞旋转，就带动周围时空旋转。有个同学曾问我关于拖曳的问题。靠近旋转黑洞有一个范围，叫做

能层，火箭进入能层就会被拖动，根本停不住。在能层里任何东西都不可能静止。必定会被转动的黑洞拖动，围着它转，这叫拖曳效应。拖曳效应是一种时空效应，是转动黑洞“拖动”周围时空，使时空跟着自己旋转的效应，由于能层中的时空被拖着转动，所以位于其中的物质被迫跟着时空一起转，不可能静止。

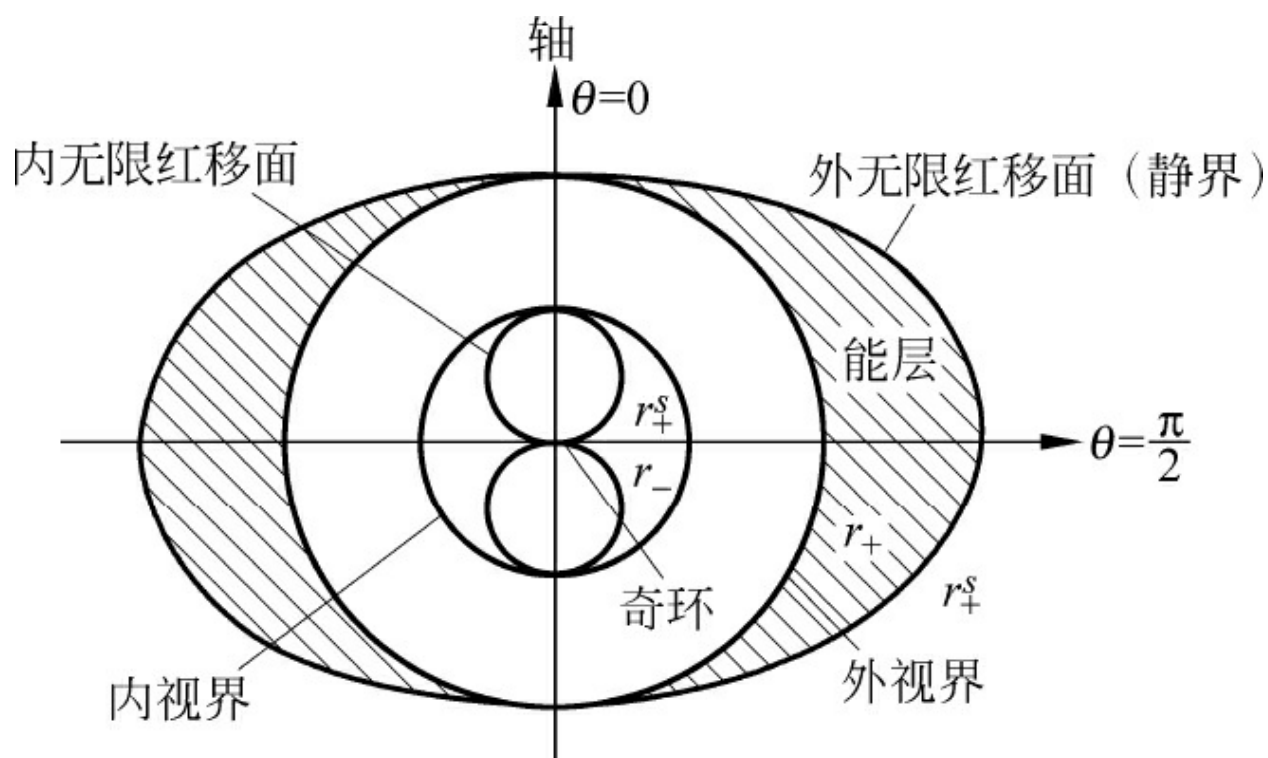


图3-14 旋转的黑洞

因为时间的限制，我们这次讲座只能先提一下黑洞形成的可能性。从今天介绍的内容看，黑洞确实是可以形成的。

大家有什么问题没有？有没有？你们应该勇敢地提问。

学生：刚才讲的牛顿力学和广义相对论，给我的感觉好像是对同一个问题从两种角度去描述，我有这种感觉。我想知道广义相对论和牛顿力学的区别到底在什么地方。是不是一个是对的，一个是错的？

答：牛顿的万有引力定律可以看作广义相对论的一个近似，这就是说在引力场很弱的情况下这两个理论是一致的。其实在地球引力场当

中，牛顿的万有引力定律已经足够精确了。所以能够检验广义相对论和牛顿万有引力定律差别的实验非常少，就只有四五个。但是呢，如果时空弯曲得很厉害，比如在中子星附近、黑洞附近，还有宇宙演化的早期，以及整个宇宙的演化，都必须用爱因斯坦的广义相对论来研究，用牛顿的理论就不行了，误差就太大了。这两个理论应该说都是正确的。只不过一个精确，一个不够精确，我们很多理论都是这样的。不够精确的理论往往仍是可用的，比如说我们原来在讨论凝聚态物理的时候，有很多东西用普通量子力学就能解决问题，甚至没有用相对论量子力学。你要用相对论量子场论来做那个东西就复杂多了。但得到的结果在有些情况下差别不大。所以很多人直接用量子力学也能解决一些问题。牛顿定律也是这样，现在我们发射卫星全部只用牛顿定律。研究白矮星也只用牛顿理论就足够了，但是研究中子星和黑洞必须用广义相对论，否则这个差异就太大了。还有研究整个宇宙的话，必须用广义相对论，用牛顿理论会有大的矛盾，大概情况是这样。还有什么？哪个同学还有问题。

学生：老师，我想问一下黑洞既然有温度，有热辐射，那么现在探测黑洞的主要方法是什么呢？

答：是这样，黑洞是有温度，但是黑洞的温度一般来说很小，太阳质量的恒星缩成半径3公里的黑洞，这么大的黑洞在望远镜中还有可能观测到，但是，它的温度只有 10^{-6}K ，比微波背景辐射的 2.7K 还要低很多，所以就无法探测到。因此直接利用霍金辐射，就是黑洞本身的热辐射来探测黑洞现在还不行。现在研究的是，当一些物质被黑洞吸积的时候，它在黑洞外旋转着往里掉，这时候会有比较激烈的效应，这样人们就有可能判断那是不是黑洞。但是令人遗憾的是，中间是颗中子星或者其他星体的话，物质往里掉，也会有类似效应。还没有找到它们之间判别性的差异。所以这个问题还解决不了。

学生：老师，我有一个问题，黑洞有没有终结？刚才老师提到黑洞

有黑体辐射，如果能量守恒依然成立的话，黑体辐射会使黑洞的能量慢慢地减少，最后黑洞是不是会消失？

答：是的，黑洞辐射以后就会变小，而且黑洞的热容量是负的，它越辐射，温度反而越上升，跟一般物体不一样，所以最后小黑洞就炸掉了。这是现在对黑洞结局的一种看法。

第三讲附录 漫谈黑洞（I）

1. 闵可夫斯基的四维时空

爱因斯坦创建狭义相对论的最初几篇论文，并没有引用四维时空的概念，这一概念是他大学时代的数学老师闵可夫斯基引进的。闵可夫斯基在为自己学生的成就感到高兴的同时，认识到如果把时间与空间看做一个整体，看做四维时空，则相对论的数学形式可以更为简洁美观。于是他引入了四维时空的概念，粗略地说，就是把时间看做“第四维空间”。

人们早已知道，空间是三维的。如果引进笛卡儿直角坐标，三维空间中两点之间的距离 dl 可以写为

$$dl^2 = dx^2 + dy^2 + dz^2 \quad (3.5)$$

如果用球坐标表示上式，就变成

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \quad (3.6)$$

闵可夫斯基把时间看做第四维空间后，构建出一个四维时空。假如仍然采用直角坐标，则两点之间的距离 ds 可以写作

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \quad (3.7)$$

假如用球坐标表示，则为

$$ds^2 = -c^2 dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \quad (3.8)$$

其中时间 t 的前面乘以光速 c ，是为了使时间项的单位与空间项一致。闵可夫斯基考虑“光速不变原理”后，正确地认识到上式中时间项和空间项之间应该差一个负号。

用式（3.7）或式（3.8）表达的四维时空，后来就称为“闵可夫斯基时空”。由于时间项与空间项之间差一个负号，所以在几何学中，闵可

夫斯基时空的几何不属于欧几里得几何，而属于伪欧几里得几何。

闵可夫斯基用自己的四维时空概念重新表述了爱因斯坦的相对论。爱因斯坦非常赞赏自己老师的这一杰作，并对他开玩笑说：您这样一改，我都看不懂自己的相对论了。闵可夫斯基的四维时空理论，为爱因斯坦后来构建广义相对论铺垫了第一块基石。

广义相对论认为，物质的存在会造成时空弯曲。不存在物质时，四维时空就是式（3.7）或式（3.8）所示的闵可夫斯基时空，当存在物质时，四维时空就会变弯，式（3.7）与式（3.8）就会有所变化。

2. 史瓦西时空——球对称的弯曲时空

德国数学、天文学家史瓦兹希尔德，算出了广义相对论的第一个严格解——史瓦西解。该解表达了当时空中存在一个不随时间变化的球对称物体时，时空弯曲的情况。这时式（3.8）将变成

$$ds^2 = - \left(1 - \frac{2GM}{c^2 r} \right) c^2 dt^2 + \left(1 - \frac{2GM}{c^2 r} \right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \quad (3.9)$$

式中M是物体的质量，G是万有引力常数，c是真空中光速。我们看到上式中，时间项与空间项之间仍然差一个负号。此式与式（3.8）的区别在于多了形如 $\left(1 - \frac{2GM}{c^2 r} \right)$ 的因子，这正是时空弯曲的表现。

不难看出，当M=0时，式（3.9）回到式（3.8），这表明物质消失时，弯曲时空恢复为平直时空。当 $r \rightarrow \infty$ 时，式（3.9）也回到式（3.8），它表明，在无穷远处质量的影响减弱，时空逐渐变得平直。

数学、物理学家们在对式（3.9）所示的弯曲时空进行研究时，发现此时空在

$$r = 0 \quad (3.10)$$

处有一个奇点。在

$$r = \frac{2GM}{c^2} \quad (3.11)$$

处有一个奇面。从式(3.9)不难看出这种情况。当 $r=0$ 时， dt^2 前的系数为无穷大。当 $r = \frac{2GM}{c^2}$ 时， dr^2 前的系数将变为无穷大。这些无穷大说明式(3.9)出现了奇异性。

进一步的研究表明，球心($r=0$)处的奇异是真奇异，时空曲率在那里发散(无穷大)，而且这一发散不能通过坐标变换来消除。我们称这种奇异性为内禀奇异性，称 $r=0$ 处为内禀奇点(或简称奇点)。

而 $r = \frac{2GM}{c^2}$ 处的奇异性是假奇异，换一个坐标系看(例如自由下落坐标系)，式(3.9)在此处的奇异性就会消失，而且时空曲率在那里正常，并不发散。人们称这种奇异性为坐标奇异性。

不过，后来的研究表明，球面 $r = \frac{2GM}{c^2}$ 处的奇异性虽然是假奇异，却有物理意义。这一球面恰是黑洞的表面，不难看出它恰是式(3.3)给出的暗星表面。

3. 时空坐标互换

再来看式(3.9)，它与式(3.8)有一点类似，时间项的前面是负号，空间项的前面是正号。我们把这一正负号差别，看做时间与空间的差别。

你们想一想，上述结论仅在

$$r > \frac{2GM}{c^2} \quad (3.12)$$

时正确，也就是说只在黑洞外部正确。当

$$r < \frac{2GM}{c^2} \quad (3.13)$$

时，式（3.9）的括号中的项将变成负值，这时 dt^2 项的前面成了正号，而 dr^2 项的前面成了负号。这一正负号的改变表明，在黑洞内部， t 变成了空间坐标，而 r 变成了时间坐标，这就是“时空坐标互换”。

所以，在黑洞内部， r =常数的曲面不再是球面，而成了“等时面”，即同一时刻的“同时面”。由于时间有方向，只能向一个方向流动，因此“等 r 面”成了单向膜。又由于黑洞内部的时间方向指向 $r=0$ ，所以任何物体都只能穿过单向膜往 $r=0$ 处跑。于是，史瓦西黑洞内部成了“单向膜区”，任何落入黑洞的物体都不能停留，都必须“与时俱进”，奔向 $r=0$ 处的奇点。也就是说，黑洞的内部除去 $r=0$ 处之外，全部是真空，没有任何物质存在。而 $r=0$ ，现在不再是球心，而是时间的“终点”。不过，这个“终点”本身并不属于时空，可以看做时空中挖掉的一个点。

需要说明一下，广义相对论并不排斥有“白洞”存在。白洞内部也是单向膜区，只不过时间方向向外， $r=0$ 处成了时间开始的地方。

黑洞是任何东西都可以掉进去，但任何东西都跑不出来的“星体”。而白洞则是不断往外喷东西，但任何东西都掉不进去的“星体”。

4. 视界与无限红移面

在第二讲中曾经谈到，广义相对论预言弯曲时空中的钟会变慢，并且导致那里的光源发出的光会发生红移，这种红移称为引力红移（即时空弯曲造成的红移），实验观测支持了这一结论。

为了探讨黑洞的性质，我们来研究一下引力红移效应。广义相对论认为，一个球对称星体（例如太阳）造成的弯曲时空中，时钟变慢由下式决定

$$\Delta t = \frac{\Delta \tau}{\sqrt{1 - \frac{2GM}{c^2 r}}} \quad (3.14)$$

式中 M 为星体质量， G 为万有引力常数， c 为真空中的光速。 τ 为静止在

星体（太阳）附近的弯曲时空中的钟走的时间， t 为无穷远处（那里时空平直。相对于太阳，地球就可看做是无穷远）观测者的钟所走的时间。由于公式根号中的因子小于1，所以 $dt > d\tau$ 。这表明太阳表面的钟走1秒时间，地球处的钟走的时间 $d\tau$ 将多于1秒，因此在地球上的观测者看来，太阳表面的钟变慢了。相应的引力红移公式为

$$\nu = \nu_0 \sqrt{1 - \frac{2GM}{c^2 r}} \quad (3.15)$$

式中 ν 为地球观测者拍到的太阳光谱线的频率， ν_0 则为地球观测者在地球实验室中拍到的同一种元素的同根光谱线的频率。从上式看，显然 $\nu < \nu_0$ ，所以在地球上的人看来，从太阳来的光线的频率减小了，即波长增大了，发生了红移。

我们在第二讲中已经谈到，实验观测支持了上述结论。现在我们来，当星体不是太阳，而是黑洞，会发生什么情况。

我们从黑洞表面直到观测者所在的位置，放置一系列钟和光源。对于靠近黑洞表面的钟和光源，由于

$$r \rightarrow r_g = 2GM/c^2 \quad (3.16)$$

式（3.14）与式（3.15）中根号内的因子趋于零。于是我们看到，位于黑洞表面附近的钟，即使 $d\tau$ 取很小的值，都会有 $dt \rightarrow \infty$ 。所以，在远方观测者看来，黑洞表面处的钟完全不走了。我们从式（3.15）则看到，不管 ν_0 取什么值，都会有 $\nu \rightarrow 0$ ，即波长 $\lambda | \lambda \rightarrow \infty |$ ，光谱线发生无限红移。因此，我们称黑洞的表面为“无限红移面”。

由于黑洞里面的任何东西都跑不出来，外部观测者得不到来自黑洞内部的任何信息，因此黑洞的表面是外部观测者能得到信息的区域的边界，所以黑洞的表面被称为“事件视界”，简称“视界”。

第四讲 霍金与黑洞



绘画：张京

霍金（图4-1）在1985年来过北京师范大学，他第一次来中国就只访问了位于合肥的中国科技大学和北京师范大学两个地方。他在北师大大的敬文讲堂（当时叫五百座教室）发表了演讲，并与北师大研究相对论的老师 and 研究生进行了学术交流。交流之余他提出，想游览长城，我们的几位研究生把他连轮椅一起抬上了长城。

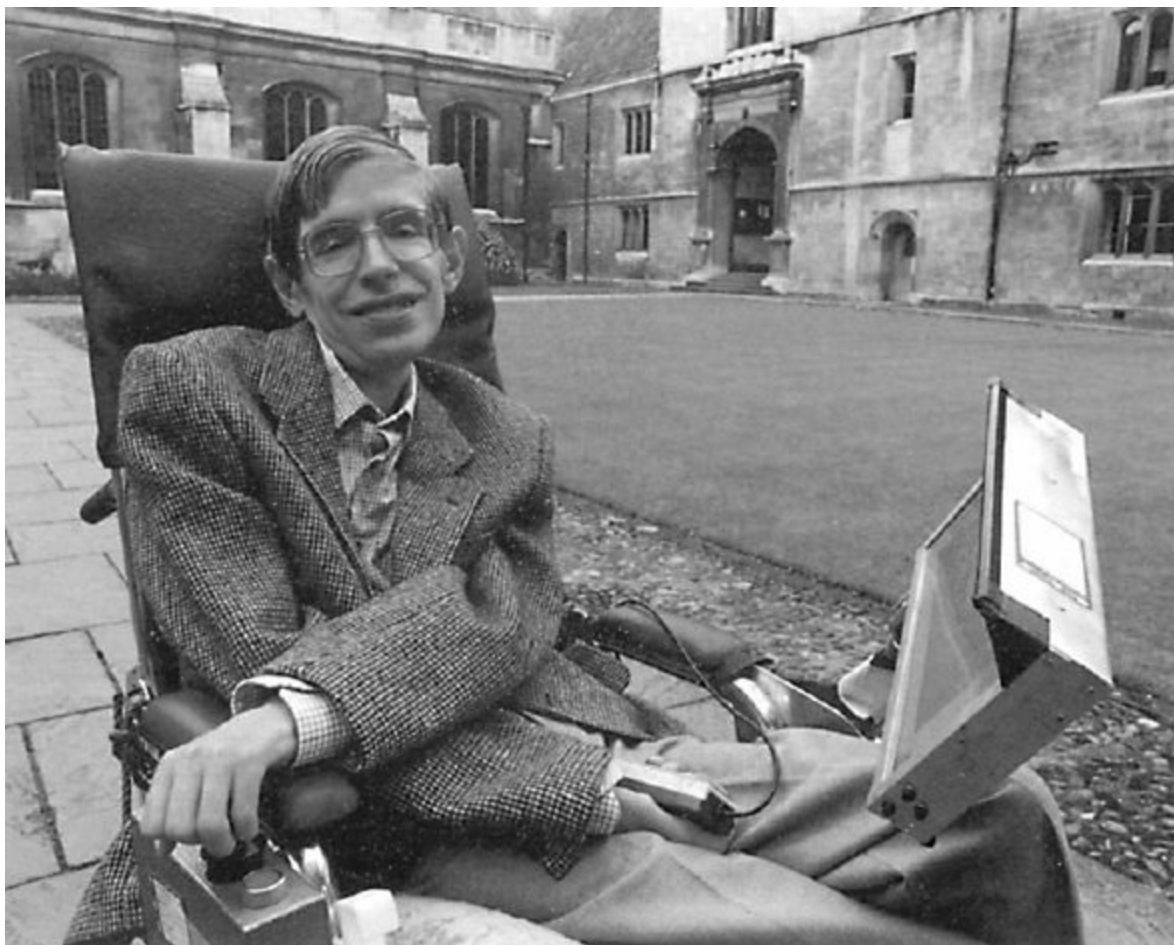


图4-1 霍金在剑桥大学校园

1. 霍金

绰号“爱因斯坦”

现在分几个部分来介绍霍金在学术上的贡献。首先让我们了解一下我们的这位明星——霍金（图4-1）。有人说他是当代的爱因斯坦，这

个提法还是可以的。霍金1942年1月8日出生于牛津，为什么他出生于牛津呢？并不是因为他家在牛津，而是因为当时二战正在进行，英国和德国达成一个默契，就是英国的飞机不炸德国的哥丁根和海德堡，德国的飞机不炸牛津和剑桥，双方都不去炸对方的文化中心，所以他母亲就到牛津生的他，那个地方比较安全一点，可以住几天。

他出生这一天，正好是伽利略逝世300周年的那一天，他经常跟人谈到这一点，意思就是你们看看，我像不像个再生的伽利略？但是他也跟别人说，其实那天出生的孩子有20万。霍金的父母都是牛津大学毕业的，父亲是学生物医学的，母亲是学文秘的。由于家里不是很有钱，他小时候上不起那种私立的很昂贵的学校，只是在一个中等偏上的学校读书。当时英国的教育制度很严格，把学生每个年级都分成A、B、C三个班，功课最好的在A班，差一点的在B班，再差的在C班。每一年要进行一次调整，A班的二十名以下的学生要降到B班，B班的前二十名要升到A班，然后B班和C班也进行这种交换，所以学生压力都非常大。霍金说第一学期他考了第二十四名，第二学期考了第二十三名，幸亏他们还有一个第三学期，考了第十八名，结果没有掉下去。他说对于掉下去的那部分学生，打击实在是太大了，他并不赞同这种制度。

在学校里，霍金的功课很一般，作业不整齐，字也写得不好，老师不怎么看好他。但是他在跟同学们聊天当中，一会儿谈一谈宇宙为什么会有红移呢？是不是光子在路上走得疲劳啦，然后就变红？一会儿又谈创造宇宙是否需要上帝帮忙啊？经常谈论这样的问题，所以同学们都比较看好他，给他取了个外号叫爱因斯坦。

牛津：人生的转折点

考大学时，霍金自己觉得考得不是很理想，但是还是考上了牛津大学。他这个人原本不喜欢物理，他说中学的物理课程简单而且枯燥，没什么意思，化学就有意思多了，为什么呢？因为化学课有时候会出现一些意想不到的事情，比如爆炸啦，着火啦之类的，所以就很有意思。一

直到中学的最后两年，受到一位老师的影响，他开始觉得物理还是挺有趣的，对整个宇宙都有所描述，对基本粒子也有描述，于是转而考了物理系，学习物理。

去牛津大学的时候正好赶上教改，英国也是在教改，只在刚进学校的时候考一次试，然后就不考了。他们的本科是三年，最后毕业的那一年再集中考一次试，在四天之内上下午连续考，把所有课全部考一遍。但是中间没有人管你，所以当时他们都放得很松。霍金回忆，他当时一天学习的时间平均不到一个小时。老师讲课也搞教改。老师来了，跟他们讲，现在讲电磁学，你们翻到第十章，你们回去自己看，后边有十三道题，过两个星期你们把作业交上来，于是老师就下课了，这课就算上完了。下课后同学们就开始做题，别看只有十三道，这十三道很难，他的同宿舍同学都只做出了一道、两道的，其余的题都没做出来。到了最后那天，第二天就该上课交作业了，霍金才想起来作业还没做，于是他没去听该上的课，赶紧补作业。其他几个人就想，这小子这时候才想起做作业来，等着看他的笑话吧。到了中午的时候，那几个同学上完课回来，问他说，你的作业做得怎么样啊？他说：这些题确实不太好做，我没做完，只做了十道。可见当时他还是比较拔尖的。

霍金上学的时候本来不用功，但是最后快毕业的时候出了一个事情：就是他有一次系鞋带时突然发现自己的手不好使了，于是上医院检查。英国的医学还是比较发达的，很快就判明了，他得的病叫进行性肌肉萎缩，不治之症，他当时才二十几岁，得这么一个病，很快就要不行了。医生也很坦率，告诉他说：哎呀，年轻人怎么得这种病了，吃点好的吧，不行了。他刚知道病情的时候，情绪一下就跌落到极点，买了些啤酒，成天在屋子里喝闷酒。他想自己大概快完了，也就两三年的事了。

过了一段时间，霍金发现一时半会儿还死不了。另外他有个女朋友，那个女朋友是牛津大学哲学系的学生，她坚持还要跟他好，说你生

病也没关系，也要跟你好。霍金一想，他还要结婚，还要养家，不能就这么混，于是他开始努力了。经过一段时间的用功，霍金发现自己还挺喜欢学习，也挺适合搞研究的，于是他就慢慢钻进去了。这次生病是霍金一生其中的一个转折点，从不用功转为用功。开始去钻研物理。

剑桥：霍伊尔与西阿玛

到了大学毕业的时候，最后这四天的考试还是很厉害的。他当时神经衰弱，考得不太满意。他们屋子里有四位同学，都还想继续上研究生，考完以后，有三个人觉得考得不行，包括霍金，只有一个人觉得考得不错，最后就是觉得考得不错的那个人没考上，这三个感觉考得不行的倒还都通过了笔试。口试的时候老师问霍金，你是留在牛津还是去剑桥？牛津和剑桥大概可以交换研究生。霍金说，你们要给我一等成绩我就去剑桥，你们要给我二等我就留在牛津，结果人家给了他一等，让他去了剑桥。

霍金到剑桥希望搞天体物理。其实他在大学的时候，原本对粒子物理有兴趣。但在六十年代的时候，粒子物理跟现在有很大的差异。那时候搞粒子物理的研究，确实就像霍金说的那样，就跟搞植物学分类相似。只研究粒子的对称性和分类，看不出什么有前景的东西来。那个领域里弱相互作用的方程没有，强相互作用的方程也没有，他觉得这太没意思了！而且发现的基本粒子越来越多，但是规律找不着，对此，他逐渐失去了兴趣。他想，起码宇宙学里还有一个爱因斯坦的相对论可以用，研究的内容还比较有趣，于是他就想改学天体物理。

最初霍金去剑桥，目的是要投奔霍伊尔。霍伊尔是著名的天体物理学家，他提出一个稳恒态宇宙模型，跟现在的大爆炸宇宙模型不一样。大爆炸宇宙就是伽莫夫的火球模型。这个模型认为宇宙刚开始起源于一个原始的核火球，然后膨胀开来，逐渐降低温度，物质密度逐渐减小，演化成了今天的宇宙。霍伊尔不同意火球模型，他认为宇宙确实是在膨胀的，但在宇宙膨胀过程中，不断有物质从真空中产生出来，所以膨胀

的时候宇宙中物质的密度基本保持稳定，跟火球模型不一样。霍伊尔不仅不同意火球模型这种观点，还讽刺火球模型的说法，说你那个模型干脆就叫大爆炸模型得了，结果这名字就用下来了，所以现在一直称这种火球模型为大爆炸模型。

起初霍金想做霍伊尔的研究生，但是霍伊尔不要他，可能是看霍金有病还是怎么的（他当时已经拄着拐棍了），反正是不要他。霍金没办法，只好找另外的教授。剑桥还有一位天体物理学家叫西阿玛。西阿玛是谁？霍金从来没听说过，可是也没法子，已经来了，霍伊尔又不要他，那就只好跟西阿玛吧。后来才发现这是一个很好的选择。西阿玛这人有一个特点，从来不主动管学生，你不是当我的研究生嘛，好，那就当我的研究生吧。但我也不管你，你不找我，我就不找你。如果你来找我，咱们俩就讨论。然后我可以给你建议，说你去找谁谁谁，或者你去看什么什么资料，看什么什么书，就这样。霍金逐渐发现，西阿玛的这种方式很好，很适合自己。

我刚开始知道西阿玛这种指导研究生的方式的时候，觉得这个导师简直不合格啊，他怎么不主动管学生呢。后来我了解到，霍金那个年龄层次的，全世界最著名的八九个相对论专家当中，有四个是西阿玛的学生，可见他带博士生的办法是对的。博士生跟硕士生不一样，不应该扶着往上走，而应该让他们自己找路往前走。

2. 相对论生涯

炮轰稳恒态宇宙模型

刚开始西阿玛没给霍金什么建议，也不管他。当时霍金还在想着霍伊尔那个模型，挺感兴趣。霍伊尔有个研究生叫纳里卡，是个印度人。霍金跑到纳里卡的办公室里，看他在干什么。国外的研究生都是一个人一个房间，或者两个人一个房间，房间大概就是9平方米，12平方米，或者更大一点。里面每人一个书桌，一个书架，还有一部电话，一块黑

板，没有其他东西。霍金进去以后就问纳里卡，“你在干嘛呢？”他说，老师布置一个东西我在做呢，霍金说我帮你算，纳里卡说当然好了。有人帮着算还不好吗？于是霍金就帮着他算，算来算去，霍金突然发现，霍伊尔这个模型大有问题，它方程里有一个系数是无穷大。霍金通过计算证明了这个东西是无穷大。你想，系数必须是有限值，不能是零和无穷大，否则系数就没用了。他发现了这个大问题，但霍伊尔还不知道。

一天霍伊尔做报告，在座的大概几十个人，讲完以后，就问，在座的有问题没有，这时霍金就拄着拐杖站起来了，说有个问题，我认为你这个系数是无穷大。霍伊尔一听，知道这个问题很大，那脸立刻拉下来了，说不是无穷大。霍金说是无穷大。霍伊尔说不是。霍金说是。霍伊尔说你怎么知道，霍金说我算过这个东西。听众开始议论，最让霍伊尔受不了的是，有几个家伙居然笑起来了。他觉得实在是太栽面子了，但是没办法。散会后，霍伊尔说，霍金是不道德的，既然知道我的论文有错，为什么不在会前向我提出，让我在会上当众出丑？与会的有些听众却认为，真正有错的是霍伊尔，为什么把不成熟的工作拿到大会上来讲？

霍金这一下子，给了稳恒态模型重重的一击。恰在那时候一些天文学家发现了微波背景辐射，这正是大爆炸模型预言的余热。从此以后，稳恒态模型就很少有人再涉及了，这下大爆炸模型完全占了上风。

奇点疑难：幸遇彭罗斯

西阿玛一生在相对论上没什么太大的贡献。他自己说过，我对相对论有两个重要贡献，第一就是把数学家彭罗斯拉过来搞相对论，第二是培养了霍金这么一个学生，他很得意。当时彭罗斯还不在牛津，是在伦敦的一个大学里工作。彭罗斯有时候来剑桥和西阿玛讨论问题，经西阿玛介绍，霍金认识了彭罗斯。

彭罗斯在研究什么呢？他在研究奇点定理，又叫奇性定理。奇点定理是怎么回事呢？广义相对论当中，黑洞里有一个奇点，曲率和密度都

是无穷大。宇宙大爆炸的时候有一个初始的奇点，大塌缩的时候有一个终结奇点，曲率和密度也都是无穷大。奇点在相对论当中是个很严重的问题，因为绝大多数时空模型都有奇点。

当时苏联有一些科学家认为，奇点其实是因为我们把对称性想得太好造成的。比如说黑洞中心有个奇点，是因为我们把黑洞想象成是在星体做精确的球对称塌缩时形成的，结果就缩成一个点。假如说不是很标准的完美球对称塌缩的话，星体中的物质就会从中间交叉错过去，就不会形成奇点了吗？他们认为奇点其实是一个偶然的現象，只不过因为我们把对称性想得太好了才出现的。

但是彭罗斯不这么想，而且彭罗斯提出一个新概念，把奇点的定义做了一个发展，他把奇点看成是时间开始或者结束的地方。白洞里边的奇点是时间开始的地方，黑洞里边的奇点是时间结束的地方，宇宙大爆炸的初始奇点是时间开始的地方，大塌缩奇点则是时间结束的地方。彭罗斯针对这个定义证明：一个合理的物理时空，如果因果性成立，有一点物质等等，在这些合理的条件之下，时空至少有一个奇点。或者说至少有一个过程，时间是有开始的，或者是有结束的，或者既有开始又有结束。这个问题可是个大问题，因为时间有没有开始和结束的问题，古代就有人讨论，但那都是哲学家和神学家的事，现在搞物理的人出来说时间有没有开始和结束。那当然很引人注目了。

霍金对这个问题很感兴趣，他的博士论文的第一部分是写稳恒态宇宙模型的错误，第二部分就是对奇点定理给出了另外的证明。当时彭罗斯已经给出了第一个证明，是针对星体塌缩成黑洞的情况。霍金又给出了另外一个证明，是针对大爆炸宇宙的初始情况。

霍金自己的第一个工作就是对奇点定理做出了贡献。他的第二个工作是黑洞面积定理，他认为黑洞的表面积随着时间前进只能增加不能减少。第三个重要工作，也是他一生当中最重要的工作就是证明了黑洞有热辐射，也就是霍金辐射，一会儿我还会讲。后来他还有一些成果，比

如说，他对时空隧道和时间机器有一些想法，为了解决宇宙奇点困难他还提出了虚时间和无边界宇宙的概念，但是我觉得他后来的这些工作都不如他青年时代的那几个工作更可靠，更有价值。

现在我们就来看他对黑洞研究的贡献。我上一讲已经介绍了黑洞的一些知识。

宇宙监督与无毛定理

但是重点只讲了简单的球对称黑洞，简单的东西容易研究，但是也有缺点，它提供给我们的知识太少。1963年，有一个叫克尔的人，求出了一个旋转星体外部的时空弯曲情况，这个解很难求，他解出来以后，很多人没有注意。他在《物理评论快报》上登了很短的一段文章，说这是爱因斯坦方程的一个解。你如果不信代进去试一试。的确是一个解，但是他怎么求出来的呢，有些人还是看不懂。

这个解有一个特点，它中心有个奇环，不是奇点。球对称的黑洞里面有个奇点，转动黑洞里边有个奇环（图4-2），黑洞的表面叫做视界，球对称黑洞的视界是球面。视界就像衣服一样把奇点包在里面，外面的人看不见它，因为视界里面的信息都出不来。转动的黑洞也有视界，是一个椭球面，它包着奇环。所以我们看不见奇环。但如果它旋转得非常厉害，最后这些视界会消失，奇环就裸露出来了。奇环一露出来，就会对时空的因果性造成破坏，所以这种情况是不应该出现的。但是研究表明，当黑洞转得非常快的时候还是会出现。彭罗斯就提出一个“宇宙监督假设”，说存在一位宇宙监督，他禁止裸奇异（裸奇点或裸奇环）的出现。

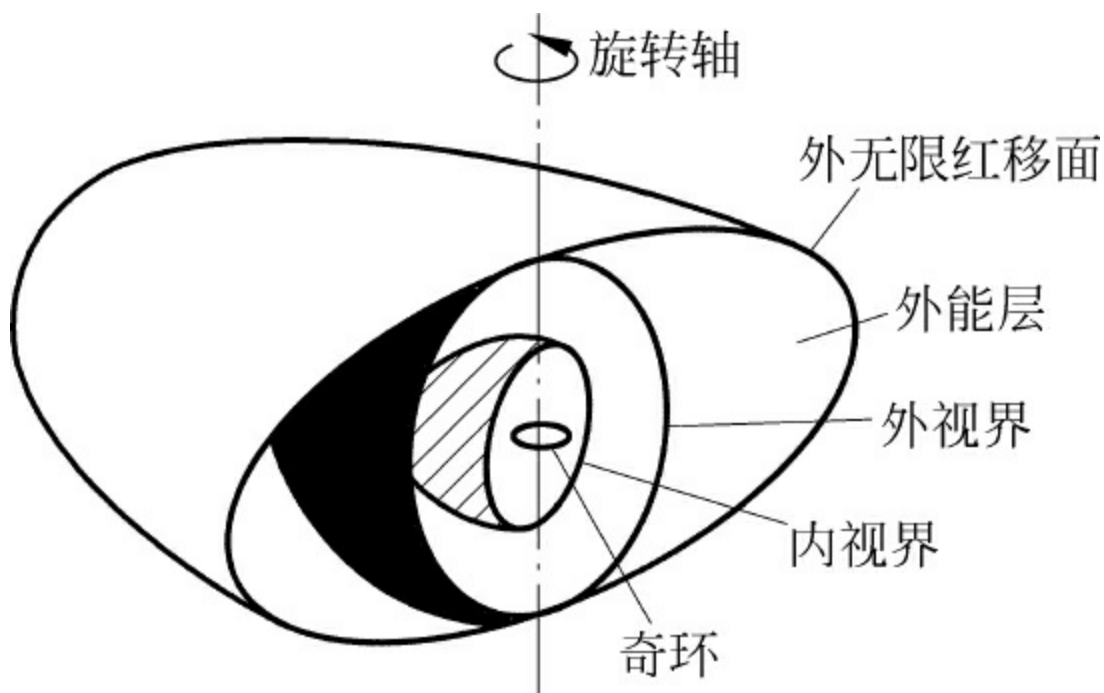


图4-2 克尔黑洞的奇环

你看，这句话等于什么也没说，就好像“自然害怕真空”一样。提出这个假设是跟他们的文化传统有关系的，我们中国人不会想出一个“宇宙监督假设”来。这是因为欧洲人是继承古希腊、古罗马文化的，在古罗马的时候，城市里有一种官员，这种官员的职责就是不准人不穿衣服在街上走，是监督官。彭罗斯认为宇宙也应该有一个监督官，不允许奇点和奇环不穿衣服露在外面，这衣服就是黑洞的视界，所以叫“宇宙监督假设”。

那么黑洞外部的人对黑洞里边能够了解什么呢？只能了解到三个信息，一个是黑洞的总质量，一个是总电荷，一个是总角动量，其他的東西都不知道，所以有人提出“无毛定理”。毛就是信息，无毛就是没有信息。但黑洞并非完全不泄漏出信息，其实还是露出了三根毛，如果我们中国人的话，肯定就叫“三毛定理”了，因为咱们有“三毛流浪记”的故事。他们就叫“无毛定理”。掉入黑洞的物质的信息都藏在黑洞里边。

3. 最伟大的发现

面积定理的启示：黑洞热吗？

总之，霍金、彭罗斯的奇点定理指出，一个合理的物理时空一定有时间开始或者结束。

我们现在来看霍金的第二个贡献：面积定理。霍金用微分几何证明了黑洞的表面积随着时间只能增大不能减小。

当时美国有一个二十几岁的研究生叫贝肯斯坦，他觉得黑洞的表面积只能增加不能减小，很像物理学当中的“熵”啊！黑洞的表面积会不会是熵啊！于是他在导师惠勒的支持下，提出黑洞的表面积可能是熵，而且他得到了一个公式，这个公式叫贝肯斯坦公式

$$dM = \frac{\kappa}{8\pi} dA + \Omega dJ + V dQ \quad (4.1)$$

这个公式大家看，很像热力学第一定律

$$dU = T dS - P dV \quad (4.2)$$

这里U是一个系统的内能，T是温度，S是熵，P是压强，V是体积，TdS是吸收的热量，PdV是对外做的功，这个你们都很熟悉。对于转动刚体

$$dU = T dS + \Omega dJ + V dQ \quad (4.3)$$

可能大家不太熟，这U是内能，TdS是热量， Ω 是转动角速度，J是角动量，V是静电势，Q是电荷。现在来看贝肯斯坦得到的黑洞的公式

(4.1)，左边dM是黑洞的质量，大家知道 Mc^2 是能量，但在选用自然单位制之后，c是等于1的，所以这个dM就是dU，等式右边后两项，非常像和功有关系的项，第一项很像黑洞的热量，其中A是黑洞的表面积，而这个 κ 叫做黑洞的表面引力。粗略地说，就是黑洞表面上如果有个质点的话 κ 就是单位质量的质点所受到的引力，叫表面引力。这个式子很像转动刚体的热力学公式，A处在熵的位置， κ 处在温度的位置。从这个公式看，黑洞不仅有熵还有温度。

争论：真热还是假热？

霍金对贝肯斯坦这个工作很不以为然，觉得贝肯斯坦完全曲解了自己的意思：我的面积定理是用微分几何和广义相对论证出来的，根本没有用到热力学和统计物理，怎么会有热呢，不可能有热。而且，一旦黑洞有温度，就应该有热辐射。黑洞是个只进不出的天体，怎么可能辐射出东西来呢？所以他在1973年的一次暑期学术研讨会上，就跟另外两个专家（卡特和巴丁），三个人合写了一篇论文，用严格的微分几何重新推导了贝肯斯坦的公式，说这个公式本身并没有错，但是它不是真正的热力学公式。它里面的 κ 像温度但不是温度，黑洞面积像熵但不是熵。于是他们就提出了“黑洞力学”的四个定律，跟普通热力学作比较。但还是强调这不是“热力学”，而是“力学”。

表4-1 黑洞力学与普通热力学的比较

	普通热力学	黑 洞 力 学
第零定律	处于热平衡的物体,具有均匀温度 T	稳态黑洞的表面上, κ 是常数
第一定律	$dU = TdS + \Omega dJ + VdQ$	$dM = \frac{\kappa}{8\pi}dA + \Omega dJ + VdQ$
第二定律	$dS \geq 0$	$dA \geq 0$
第三定律	不能通过有限次操作,使 T 降到零	不能通过有限次操作,使 κ 降到零

霍金辐射的发现

可是霍金后来又想：万一贝肯斯坦是对的呢？他又倒过来想了。如果贝肯斯坦是对的，那么黑洞就真有温度，就应该有热辐射射出。于是他又经过半年多的努力，在1974年终于证明了黑洞确实有热辐射，黑洞的温度是真温度。那个 κ 反映的是真温度，黑洞表面积 A 确实是熵，就是黑洞熵。严格证明黑洞有热辐射，是霍金一生中最卓越的成就。后来人们就把黑洞热辐射称为霍金辐射。这件工作做出来以后，他的老师西阿玛就说：霍金毫无疑问是20世纪最伟大的物理学家之一。

霍金辐射刚开始提出来的时候，很多人都接受不了，有些人觉得他是胡说。他在英国剑桥大学第一次报告这一工作的时候，刚刚讲完，主

主持会议的那位教授就说：刚才霍金博士给我们做了一个精彩的演讲，很有意思，但都是胡扯，然后就上厕所去了。回来后大家又讨论了一番。最后表明霍金的想法是对的。

这是怎么回事呢？前面讲过，黑洞里边时间箭头朝里，任何物质和辐射只能往里掉，不能跑出来。从经典的广义相对论考虑，确实不可能有热辐射从黑洞射出。现在霍金考虑量子效应，他用弯曲时空背景下的量子场论来研究黑洞附近的情况。

黑洞附近的真空涨落

我们知道，在平直时空中，真空是不空的，不断有虚的正反粒子对产生，产生出来又湮灭，产生出来又湮灭，这叫真空涨落。正反粒子对中，一个粒子是正能，另一个是负能，符合能量守恒，它们产生出来很快就又消失了。由于 Δt 和 ΔE 的测不准关系，

$$\Delta t \Delta E \sim \frac{h}{2} \quad (4.4)$$

在这么短的时间之内你不可能测到负能粒子。你在 Δt 的时间里一测，就有相当于 $\Delta E \sim \frac{h}{2\Delta t}$ 的不确定的能量出现，把负能粒子掩盖掉。所以测不出负能粒子的存在。这种真空涨落已经被许多物理实验间接证明了，所有搞量子论的专家都承认真空涨落。

霍金现在研究黑洞附近的真空涨落。他说真空涨落假如发生在黑洞附近，会有几种可能情况出现。一种可能就是两个粒子都没掉进去就复合而消失了，与平直时空情况差不多，没有什么特殊效应；另一种可能就是两个粒子都掉进去了，那也没什么效应；第三种可能是负能的粒子进去了，正能的跑出来了（图4-3）。有人会问，会不会有第四种可能：正能粒子掉进去，负能粒子跑出来。不可能！因为黑洞外边的时空就是我们普通的时空，它不允许负能粒子单独存在，只有黑洞里面的时空，才允许负能粒子单独存在，如果正能粒子掉进黑洞，负能粒子必然跟着掉进去。

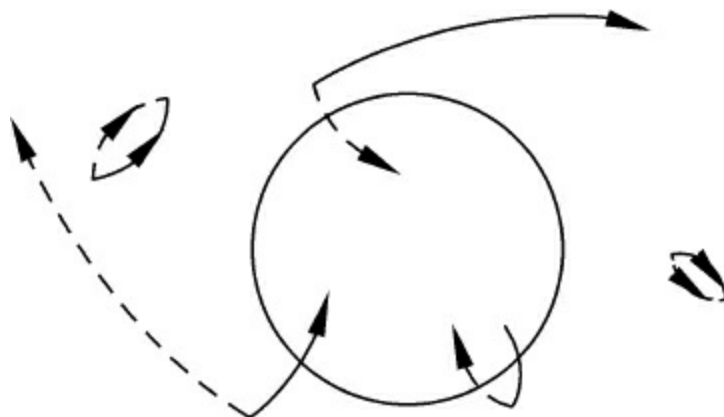


图4-3 黑洞附近的真空涨落

现在我们来第三种情况，负能粒子掉进黑洞，它顺着时间的发展落向奇点，使奇点减少一个粒子的质量。而正能粒子飞向远方。例如产生了一个正反电子对，由正能电子和负能正电子组成。其中，负能正电子落进黑洞，顺时前进落向奇点，使奇点处减少一个电子的质量，同时增加一个正电荷。而正能电子（带负电）飞向远方。霍金认为，这一过程相当于从奇点处产生一个正能电子，逆着时间前进飞到黑洞表面，被视界散射，再顺着时间方向飞向远方。对遥远的观测者来说，他收到了一个带负电的正能电子，而黑洞减少了一个电子的质量，增加了一个正电荷。霍金用弯曲时空量子场论严格证明了黑洞确实会产生这种量子辐射效应，而且射出的粒子的能谱是严格的普朗克黑体辐射谱。也就是说黑洞确实会产生热辐射。霍金对量子辐射的这种解释，既不违背能量守恒和电荷守恒，又不违背黑洞的定义，非常成功，非常合理。

黑洞的温度与熵

霍金严格证明了黑洞的温度为

$$T = \frac{\kappa}{2\pi k_B} \quad (4.5)$$

熵为

$$S = k_B \frac{A}{4} \quad (4.6)$$

其中 k_B 为玻耳兹曼常量。

对于球对称的史瓦西黑洞，其表面引力为

$$\kappa = \frac{1}{4M} \quad (4.7)$$

我们看到 κ 与黑洞质量成反比，也就是说，黑洞的温度与质量成反比。

奇妙的负比热

如果霍金辐射不断进行，黑洞就会逐渐消失。为什么呢？因为黑洞的温度与质量成反比，所以黑洞的比热是负的。一般物体的比热都是正的，如果放出热量温度就下降。但是，黑洞不一样，辐射粒子以后，质量减小，温度反而会升高，所以黑洞和外界不可能处于稳定的热平衡。即使达到热平衡，只要有一个涨落，黑洞比外界温度高一点，就会辐射粒子，辐射粒子后温度会变得更高，温差就拉大了，辐射也会变得越来越厉害，最后小黑洞就炸掉了。

如果刚开始黑洞跟外界热平衡，一个涨落使黑洞的温度低一点，那么外界的能量就流进来了。外界的能量一流进来，黑洞质量增加，温度反而降下去了。会有更多的能量往里流，这样黑洞就不断长大了。总之，黑洞与外界不可能处在稳定的热平衡当中。

另外还有一些现象，例如吸积和喷流（图4-4），这方面搞天体物理的人研究得比较多。假如有两颗恒星，一颗已经形成黑洞，另外一颗恒星的气体会被黑洞吸过去，形成吸积盘，旋转着往里掉。这些东西往里掉的时候会有很激烈的效应。物质掉进黑洞的时候，在吸积盘的垂直轴方向会产生喷流。现在喷流现象已经在天文学上看到很多了，但是中心的这个星体不是黑洞也会有喷流，所以目前吸积、喷流现象仍然不能作为黑洞存在的最终判据。

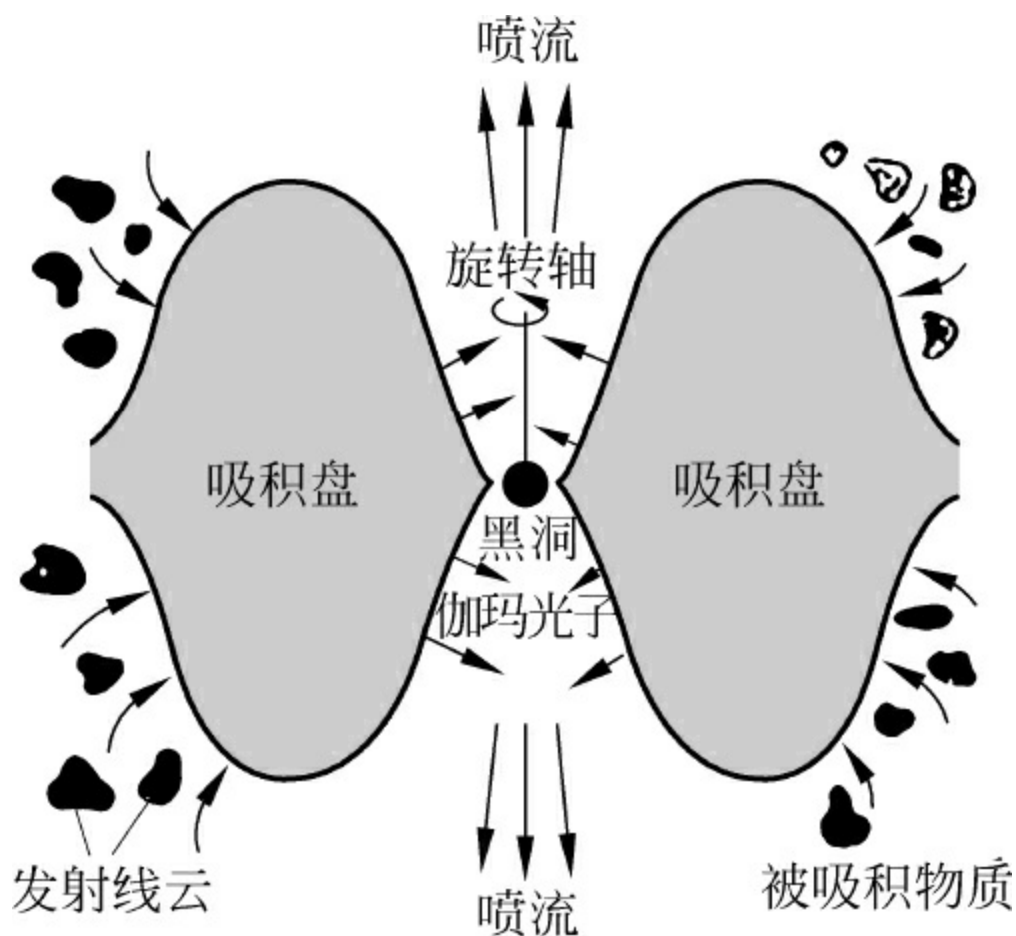


图4-4 黑洞的吸积与喷流

4. 信息疑难

信息守恒吗？

现在我们来谈一下有关霍金的另外一件事情，就是关于信息守恒的问题。我们知道，无毛定理说，东西掉进黑洞以后，外界的人就不知道它们的信息了，但是这些信息并没有从宇宙中消失，它们藏在了黑洞的内部，外界的人只能探知三个信息，就是总质量、总电荷和总角动量这三根毛。这件事情问题还不是很大。

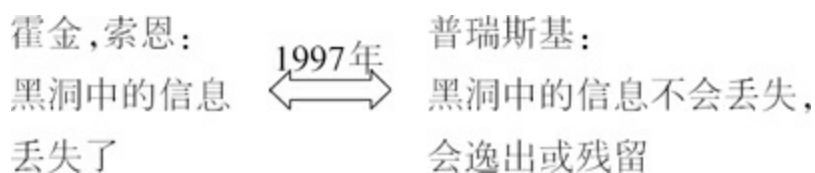
但是，认识到黑洞有霍金辐射以后，问题更大了：研究表明，黑洞往外辐射正反粒子的概率相同，电子和正电子概率相同，质子和反质子概率相同，而且完全是热辐射，辐射谱是标准黑体谱，而热辐射是几乎

不带任何信息的，所以没有信息伴随霍金辐射跑出来。而且，黑洞是越辐射温度越高，那么黑洞就越变越小，最后就爆炸消失了。这样，原来掉进黑洞的物质有大量的信息，最后辐射出来的物质基本不带信息，黑洞又消失了，那么这些信息不就从宇宙中丢失了吗？信息就不守恒了。

这件事情引起了理论物理界的争吵，搞相对论的人认为信息不守恒就不守恒吧，没什么关系。但是搞粒子物理的人可不这么认为，信息不守恒会导致概率不守恒，这样量子么正演化的规律就有问题了。整个粒子物理的基础要动摇，所以他们都认为信息应该守恒。他们猜测，可能霍金辐射不是严格的热辐射，会有一些信息带出来。要不然就是黑洞蒸发到某一个阶段会突然截止，有某种像量子效应一样的东西把霍金辐射一下截止，剩下的那些信息都作为炉渣沉在黑洞里面，不会消失。

霍金打赌

1997年，霍金和索恩（Kip Thorne），就是搞时空隧道和时间机器的那位专家，他们两个人说黑洞中的信息会丢失，粒子物理学家普瑞斯基说不会丢失。于是，他们打赌，谁输了谁给对方订一年棒球杂志。他们打赌是开玩笑的。



到了2004年7月，霍金突然宣布说：我输了，我承认信息是守恒的。而且在爱尔兰开国际相对论大会的时候，买了一堆杂志给普瑞斯基带去了。索恩说：这事不能由霍金一个人说了算，我不承认输了。普瑞斯基则说：没听懂自己是怎么赢的。虽然霍金承认输了，但是他没听懂霍金为什么输，自己为什么赢。霍金的主要意思是，原先把黑洞想象得太理想了，因为时间关系我不可能跟大家讲很多。总之，他认为，真正的黑洞并不是我们想象的那种理想的东西，信息不会丢失。

霍金：我输了 2004年 普瑞斯基：没有听懂我为什么赢了
 索恩：没有输

霍金改变态度的一个原因是，当时已经有一些人做了这方面的证明，比如说帕瑞克和维尔赛克。维尔赛克是诺贝尔物理学奖获得者，搞强相互作用的。他们做出了一个证明，证明信息是守恒的。他们很巧妙，说你看，霍金在证明热辐射的时候，考虑射出光子，射了一个光子以后，黑洞的质量不就减少了一个光子的质量吗？质量减小黑洞半径不就会减小吗？（图4-5，图4-6）但霍金没有考虑质量减小的影响。霍金确实没有考虑，其他人（包括我们自己）也没有考虑。

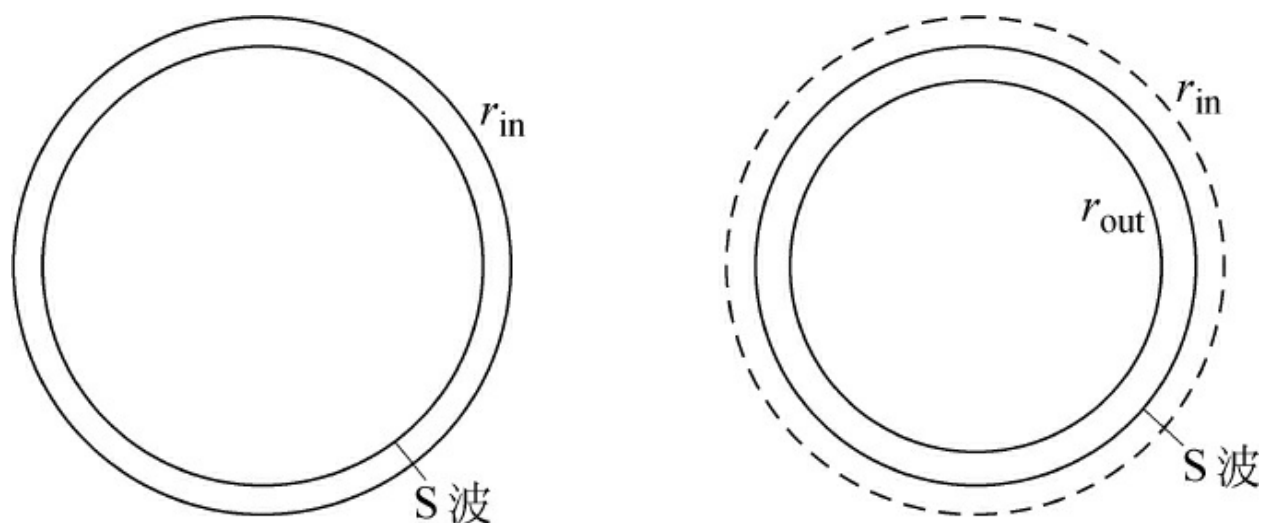


图4-5 隧穿过程：辐射使黑洞收缩。势垒在黑洞视界处，粒子以球形波（S波）形式向洞外隧穿， r_{in} 和 r_{out} 分别为射出前和射出后的视界位置



图4-6 隧穿示意图：好像粒子（图中小人）不动，势垒向内移动，从 r_{in} 移到 r_{out}

关于一个转动的黑洞辐射电子的情况，是我们这个组在刘辽先生和北大的许殿彦先生领导下首先证明的，我们证明的时候也确实没考虑这一点。我们当时觉得没考虑是完全可以的，为什么呢？大家知道，太阳质量占整个太阳系质量的98%或99%，太阳形成黑洞以后半径三公里，跑出一个电子或跑出一个光子，太阳质量能变化多少？半径能缩小多少？简直太微乎其微了。因此，所有证明黑洞辐射的人都没考虑这个问题。帕瑞克他们说，就是因为没考虑这一点，黑洞辐射才是严格的黑体辐射，信息才跑不出来。他们做了一个证明，出去一个粒子以后，真的有一点影响，那点影响就能对热谱有一点修正，这一点修正就正好把信息带出来了，于是信息就守恒了。

对信息守恒的质疑

我当时有一个博士生张靖仪，正好开始做论文，我就建议他研究这个问题，把帕瑞克的工作推广到各种黑洞，因为他们做的是最简单的球对称黑洞情况。我说我觉得他们的证明可能不对，他们暗中可能有一个假定，假定了这个过程是可逆过程。为什么呢？我的主要想法是：搞信息论的人认为信息是负熵，这点已经被很多物理学家所接受了，包括霍金本人都认为信息是负熵。

大家知道热力学第二定律的精髓就是熵不守恒。在一个真实的自然过程中，熵是会增加的，只有理想的可逆过程，熵才会守恒。所以根本没有道理说，一定要维持一个信息守恒定律，物理学当中现在没有，将来也不一定必须有这么一个定律。如果真的信息就是负熵的话，恐怕信息就是应该不守恒的。由于帕瑞克等人研究的是最简单的黑洞，模型简单，提供的信息也就少，不容易看出他们是否用了“可逆过程”这一假定。假如能够将帕瑞克的工作推广到各种复杂的黑洞，也许就容易看清楚了。

所以我建议张靖仪研究这个问题。后来他在博士生期间做了八篇论文，都是在国外杂志上刊登的，刘文彪教授和胡亚鹏、方恒忠、任军等

研究生也做了一些研究。张靖仪做到第八篇的时候，我们看出来了，帕瑞克的证明方案中确实暗含了一个假定：假定了过程是可逆过程。他们用的热量变化式是 TdS 这个式子，只有在准静态的可逆的过程中才可以用这个式子。用这个式子就等于假定了过程是可逆过程，熵当然守恒，信息自然也守恒。所以，帕瑞克等人的工作是在可逆过程的假定下证明的。而实际的自然过程是不可逆的，具有负比热的黑洞，由于不存在稳定的热平衡，它的热辐射过程肯定是不可逆的。所以帕瑞克他们的工作，虽然数学上是正确的，但并无实际的物理意义。

我觉得现在黑洞的问题要分几个方面，一个方面，确实如霍金讲的，以前把黑洞考虑得太理想化了。真实的霍金辐射有可能带出一部分信息。另一方面，如果信息确实是负熵的话，没有道理要求信息一定守恒。而且由于熵不守恒，应该推测信息是不守恒的，由此看来霍金打赌的这场争论仍然会继续下去。

我今天对霍金与黑洞的介绍就到这里。我想把霍金的一句话作为我们今天这个报告的结束：“当爱因斯坦讲上帝不掷骰子的时候，他错了，对黑洞的思索向人们提示，上帝不仅掷骰子，而且有时还把骰子掷到人们看不见的地方去了。”那个地方是什么呢，就是黑洞。到现在为止，黑洞的问题还没有完全搞清楚，还要继续研究下去。而霍金本人，由于他的成就，已经成为20世纪最伟大的物理学家之一了，这是毫无疑问的。

这一讲就到此结束。

第四讲附录 漫谈黑洞（II）

1. 最一般的黑洞

现在已经证明，自然界中可能存在的、不随时间变化的黑洞，是转轴对称的带电黑洞，称为克尔—纽曼黑洞。这种黑洞不仅具有质量 M ，还带有角动量 J 和电荷 Q 。（参见图3-14和图4-2）

研究发现，这种黑洞比球对称的史瓦西黑洞复杂得多，它的视界分为两个

$$r_{\pm} = \frac{GM}{c^2} \pm \sqrt{\left(\frac{GM}{c^2}\right)^2 - \left(\frac{J}{Mc}\right)^2 - \frac{GQ^2}{c^4}} \quad (4.8)$$

+号表示外视界，-号表示内视界，式中 G 是万有引力常数， c 是真空中光速。

为了突出上式的物理内涵，人们采用自然单位制，即令 $c=G=\hbar=1$ ，这样上式就可简化为

$$r_{\pm} = M \pm \sqrt{M^2 - a^2 - Q^2} \quad (4.9)$$

式中， $a = \frac{J}{M}$ 为单位质量的角动量。

研究还表明，这种黑洞的无限红移面与视界分开了，而且也分成两个

$$r_{\pm}^s = M \pm \sqrt{M^2 - a^2 \cos^2 \theta - Q^2} \quad (4.10)$$

其中+号表示外无限红移面，-号表示内无限红移面。在无限红移面与视界之间，夹着能层，能层里虽然是真空，但储存着能量。在 r_{+}^s 与 r_{+} 之间是外能层，在 r_{-}^s 与 r_{-} 之间是内能层。

外无限红移面像一个橘子的外皮，而内无限红移面像一个花生的外壳。从图3-14看，内、外视界似乎是两个球面，式（4.9）好像也支持这

一点。因为这种黑洞的质量 M 、角动量 J 和电荷 Q 都不变化，似乎 r_+ 和 r_- 都与角度无关。但这是一种误解，研究表明，式（4.9）所示的克尔—纽曼黑洞的内、外视界面实际上都是椭球面。这是因为我们此处用的坐标是椭球坐标，不是大家通常见到的球坐标。对于球坐标， $r=0$ 是一个点；但对于椭球坐标， $r=0$ 不是一个点，而是一个小圆盘。史瓦西黑洞的奇点，在克尔—纽曼黑洞中变成了奇环。这个奇环就是 $r=0$ 的小盘的外沿，用椭球坐标表示，奇环处在 $r=0$ 且 $\theta = \frac{\pi}{2}$ 处，即

$$\begin{cases} r = 0 \\ \theta = \frac{\pi}{2} \end{cases} \quad (4.11)$$

这真是很怪异的事情。

在位于无穷远的观测者看来，离外无限红移面越近的钟走得越慢，离那里越近的光源射过来的光红移量越大。放置在外无限红移面上的钟，干脆就不走了，放置在那里的光源发来的光，会发生无限大的红移， $\lambda \rightarrow \infty$ 。这就是无限红移面名称的由来。

穿过无限红移面进入能层的物体，将不可能静止，一定会被转动的黑洞拖着旋转，这叫拖曳效应。拖曳效应是一种“时空效应”。研究表明，能层内的物体如果转动角速度为零，所处的状态就是超光速运动状态，而超光速运动是相对论所禁止的，所以能层内的物体的角速度不可能是零，必须被转动黑洞拖动。这样看来，无限红移面是“物体可以静止”的边界，所以又称它为“静界”。

不过，克尔—纽曼黑洞的表面不是外无限红移面，而是外视界。穿过外无限红移面进入外能层的飞船，只要不进入外视界，就可以再飞出去。外能层区不是单向膜区。时空坐标互换的单向膜区位于内外两个视界之间，所以进入外视界的飞船不再能逃出来，它必须“与时俱进”，奔向并穿过内视界，进入内视界以里（ $r < r_-$ ）的时空区，那里不再是单向

膜区， r 重新成为空间坐标， t 重新成为时间坐标。所以，进入那里的飞船不会毁灭，可以永远在那里停留，但不可能再飞出来了。

早先人们认为，进入内视界以里（ $r < r_-$ ）的飞船，只要小心控制，不要碰到奇环就行。后来的研究表明，飞船根本不可能碰到奇环，奇环有一股强大的排斥力，拒绝任何物体向它靠近。

2. 极端黑洞、宇宙监督假设与无毛定理

从式（4.9）不难看出，如果不断地增加克尔—纽曼黑洞的角动量和电荷，将会有

$$a^2 + Q^2 \rightarrow M^2 \quad (4.12)$$

式（4.9）中的根号将趋于零，这表明内、外视界会相互靠近，单向膜区的厚度将变薄。当

$$a^2 + Q^2 = M^2 \quad (4.13)$$

时，式（4.9）中根号为零，内、外视界将重合

$$r_+ = r_- = M \quad (4.14)$$

单向膜区将收缩成一张厚度为零的膜，这时的黑洞称为极端黑洞。

如果再向极端黑洞输进角动量和电荷，将会有

$$a^2 + Q^2 > M^2 \quad (4.15)$$

式（4.9）中的 r_{\pm} 将成为复数，这样的几何面是不存在的，这意味着视界和单向膜区都会消失，奇环将裸露在外面。

由于奇环会破坏时空的因果性，外部观测者看到裸露出来的奇环，将使他的因果演化变得不确定。显然，应该有一条物理定律禁止奇环的裸露。因为这个问题一时难以弄清楚，彭罗斯提出一个假设——宇宙监督假设：

“存在一位宇宙监督，它禁止裸奇异（奇点或奇环）的出现。”

对于这位“宇宙监督”究竟是谁，目前还没有一致的意见。

从式（4.9）和式（4.10）容易看出，克尔-纽曼黑洞的视界和无限红移面，由三个物理量决定， M 、 J 、 Q 。事实上，黑洞外部的观测者，只能探知黑洞的这三个信息：总质量 M 、总角动量 J 和总电荷 Q 。形成黑洞和后来进入黑洞的物质的其他信息都探测不到了。黑洞是一颗忘了本的“星”，它忘记了自己原来是一颗什么样的星，忘记了它是怎样形成黑洞的，也忘记了它形成后又有哪些东西掉进去。

科学家们提出“无毛定理”，“毛”就是信息，认为黑洞形成时信息丢失了，黑洞没有毛。实际上黑洞还剩三根毛，那就是 M 、 J 、 Q 。

3. 黑洞的温度

与史瓦西黑洞一样，克尔-纽曼黑洞也有温度，有热辐射。研究表明，这种黑洞的温度和熵也由式（4.5）和式（4.6）决定。

克尔-纽曼黑洞的表面引力的表达式比较复杂，由

$$\kappa = \frac{r_+ - r_-}{2(r_+^2 + a^2)} \quad (4.16)$$

表示。

从式（4.16）可知，极端黑洞的 $\kappa=0$ ，也就是说，极端黑洞的温度是热力学温度绝对零度。所以，不少人推测“宇宙监督”就是热力学第三定律。第三定律认为：不可能通过有限次操作，把系统的温度降到热力学温度绝对零度。对于黑洞来说，就是禁止黑洞演化成极端黑洞。极端黑洞尚存一张视界膜，如果达不到极端黑洞，当然就更不可能让这层膜消失，奇环也就裸露不了。

4. 安鲁效应

在霍金提出黑洞有热辐射的前夕，安鲁（W.G.Unruh）发现，匀加速直线运动的伦德勒观测者处在热浴中。这就是说，原本一无所有的闵

氏时空，所有惯性观测者均认为是真空，但是，在其中作匀加速直线运动的观测者会发现自己周围充满了热辐射，其温度为

$$T = \frac{a}{2\pi k_{\text{B}}} \quad (4.17)$$

这个温度取决于伦德勒系的加速度 a 。

安鲁的结论是惊人的。然而，由于大部分物理工作者不熟悉广义相对论，也由于这一效应过于微弱，目前在实验中还观测不到，这一杰出的工作至今还不为世人所注意，只有少数人知道有这个已被预言但尚未观测到的效应存在。

安鲁等人认为，伦德勒观测者感受到的热效应是一种量子效应，它是由于不同参考系有不同的“真空”而造成的。按照狄拉克的思想，真空不空，有零点能存在。通常的物理学都是在平直的闵氏时空的惯性系中讨论的，所以物理学中所说的真空，通常都是指惯性系中的真空，闵氏真空的虚粒子涨落形成零点能（图4-7）。当我们在作匀加速直线运动的伦德勒系中观测时，由于伦德勒真空不是闵氏真空，它的能量零点比闵氏真空的能量零点要低（图4-8），因此，闵氏真空的零点能在伦德勒观测者看来就是高于真空零点的能量，是真实可测的能量。这种能量以最简单的形态出现，那就是具有黑体谱的热辐射状态。因此，伦德勒观测者觉得自己浸泡在热浴之中。

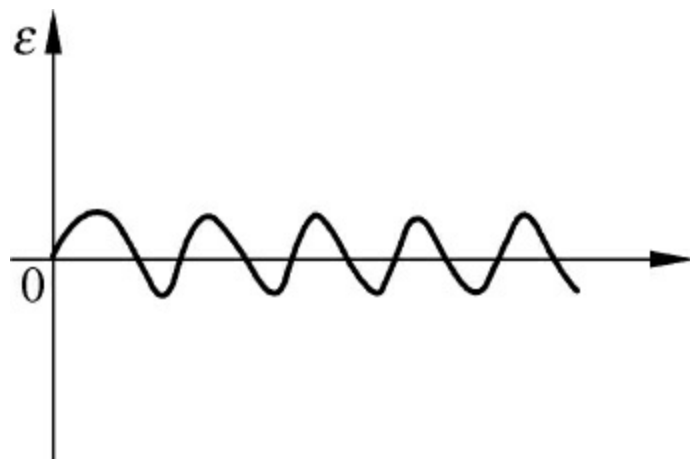


图4-7 闵氏时空零点能

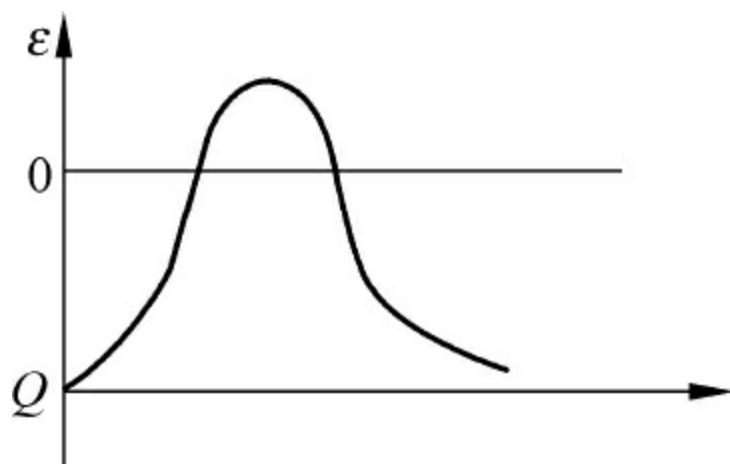
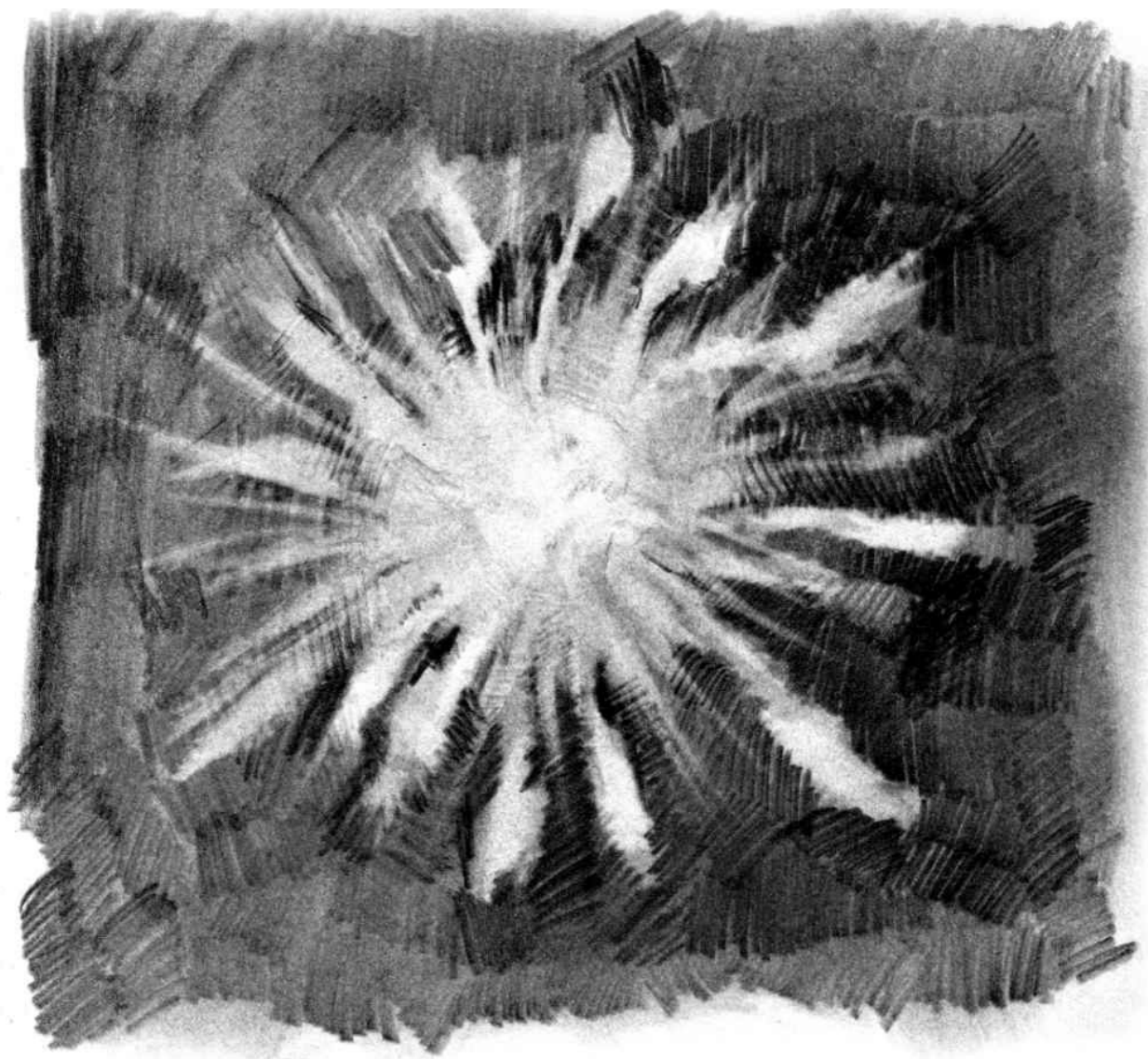


图4-8 伦德勒时空，真空能量的零点下降到Q点，闵氏真空的零点能以热能形式出现

霍金证明了黑洞有热辐射之后，安鲁很快认识到，自己发现的效应与黑洞热辐射有相同的本质。因此许多人把霍金辐射与安鲁效应一起，称为霍金—安鲁效应。

第五讲 膨胀的宇宙



绘画：张京

我们现在就来讲一个新的讲座——膨胀的宇宙。前面已经讲了四讲了，这次我们来讲宇宙。我们主要讲一下现代的宇宙模型，但是为了引进这个现代宇宙模型，我们要先讲一下宇宙的结构。

什么叫宇宙呢，汉朝的时候淮南王刘安招集一批门客写了一部书，叫做《淮南子》。后来有一位叫高诱的人，在《淮南子》的《原道篇》里边加了一个注，对宇宙下了个定义说：“四方上下曰宇，往古来今曰宙。”这就是说，宇就是空间，宙就是时间。不过，我们今天说宇宙的时候，是把时间、空间和物质总括到一起称为宇宙的。

那么我们就来看一下宇宙的结构。像温总理讲的，一个民族总要有的一些仰望星空的人。当然那是指考虑民族远景和大的发展方向的人。现在我们来看一下真正的星空，再从能看见的星空入手，讲解一下我们的宇宙。

1. 浩瀚的星空

猎户当空，三星高照

大家看，图5-1是冬夜的星空，基本上是春节前看到的星空。晚上八九点钟往南看到的天空就是这样。我们看到天空繁星万点，最明显的是中间这个四边形，这是猎户座，中国名称是二十八宿里的参宿。希腊的名字是猎户座，中间横着三颗星，是参宿一、参宿二和参宿三。我们通常说“三星高照，春节来到”，就指这三颗星。



图5-1 冬夜的星空

猎户座的左下方是大犬座，大犬座的 α 星就是我们肉眼所能看见的、除太阳以外最亮的恒星——天狼星。我们上次讲黑洞的时候曾经提到过天狼星。天狼星有一颗伴星是颗白矮星，是人类发现的第一颗白矮星。天狼星的左下方是弧矢星。

中国古代认为天狼星代表侵略，所谓“弧矢射天狼”，就是反击侵略。我上次讲黑洞的时候曾经提到苏东坡的诗：“会挽雕弓如满月，西北望，射天狼。”为什么“西北望，射天狼”呢？一个原因是，西夏是北

宋当时的主要敌人，在中原的西北方向。另外一个原因是天狼星出现在弧矢星的西北。有一个网友说天狼星从来不出现在天空的西北，他讲的是对的，天狼星一直是在南面的天空，不过它是在弧矢星的西北方。

猎户座的右上方是金牛座，金牛座里有我们所说的蟹状星云，星云中心有一颗中子星。

银河缥缈，繁星万亿

我们看，夜空中有一个淡淡的白条，那就是我们的银河。因为现在城市里的空气污染很厉害，而且光线也太强，所以我们经常看不到很美的夜空。如果你们到郊外的话，就会看到这样美丽的星空。这万点繁星其实都是银河系里边的恒星，银河系之外的恒星，我们用肉眼是看不见的，我们最多只能看见像仙女座星系这样少数几个河外星系。它们是像银河系一样的星系，肉眼看来是模糊的斑点，有点像恒星，用望远镜仔细看，它们都是与我们的银河系类似的星系，每一个都由上千亿颗恒星组成。

我们的银河系直径是十万光年，就是说光从银河系的一端走到另一端需要十万年。它由一千亿到两千亿颗恒星组成，按今天的说法，大概是两千亿颗恒星。银河系2.5亿年自转一周。太阳系不处在银河系的中心，位于比较靠边一点的地方，以每秒250公里的速度围绕银河系的中心旋转。

这两千亿颗恒星组成了几百亿个太阳系。为什么不是组成两千亿个太阳系呢？这是因为大部分太阳系都有两个以上的太阳。真正像我们的太阳系这样，只有一颗恒星的太阳系是比较少的。

星移斗转，北极定向

作为例子给大家看一个由好多个太阳组成的太阳系。大家看这张图（图5-2），这张图下边是大熊星座，上边是小熊星座。大熊星座就是北斗七星，把北斗七星右端的两颗星连起来，再把它延长五倍就是北极星。小熊星座最亮的那颗星，就是北极星。大家都知道，在郊外要是迷

路了，顺着北斗就能找北极星。

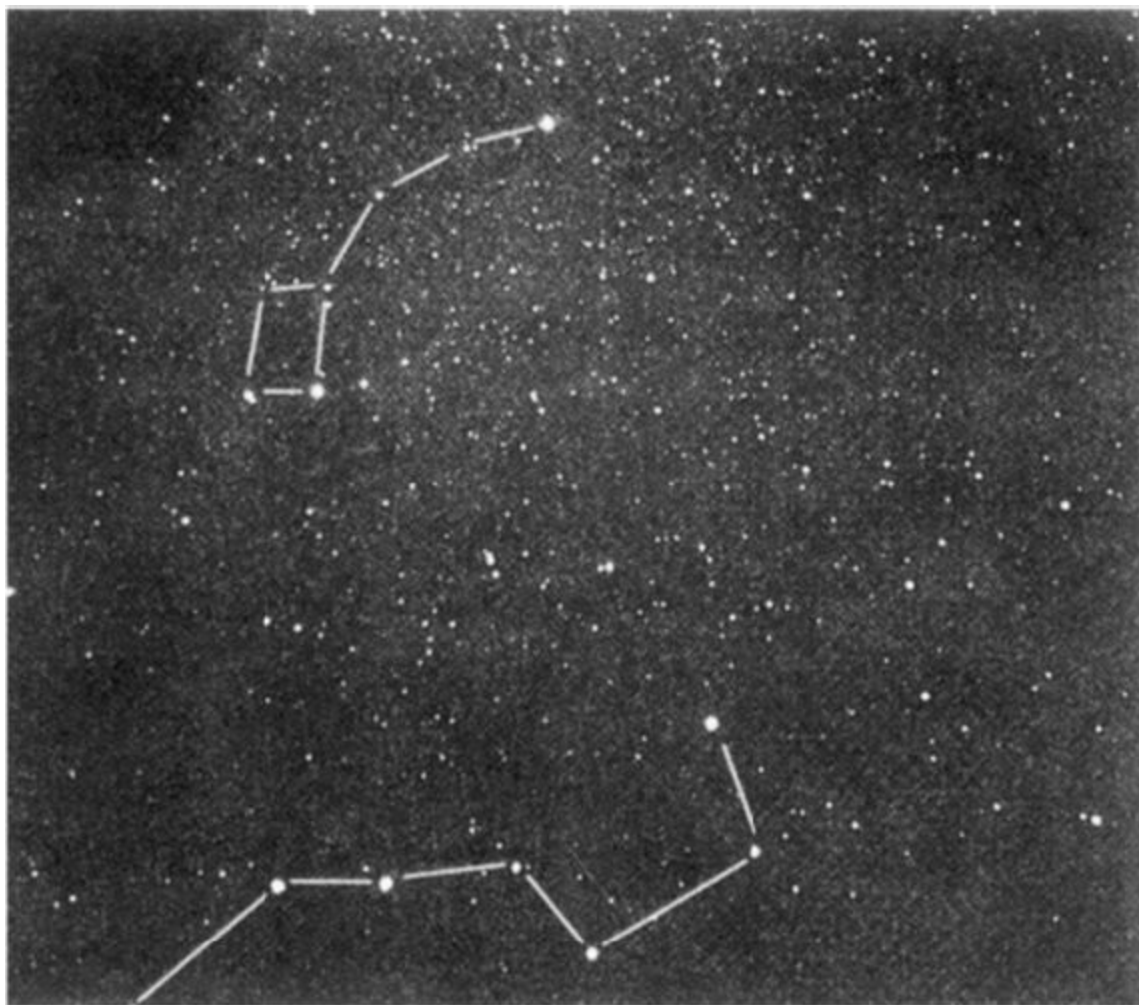


图5-2 大熊星座与小熊星座

天上的群星，包括北斗星都要围绕着北天极转。北极星就在北天极附近，所以天上的群星似乎都在围绕北极星转。可是北斗有的时候会转到下边来，转到下边的时候可能被高山挡住，就可能看不见北斗。但那时候仙后座就升起来了，仙后座有五颗亮星，组成字母W，W的缺口也指着北极星。所以你只要认识星，就可以找到北极星。当然，从来不认识星的人，到了迷路的时候，看哪儿都像北斗。不过你静下心来仔细观察，还是能看出哪些星比较亮，北斗七星那几颗确实比周围的星亮，你还是能够认出北斗的。

恒星的命名

我们今天要讲的是北斗七星中的一颗。图5-3是北斗七星，它们是大熊星座里面的星。最亮的那颗星是大熊星座 α ，次亮的叫大熊座 β ，这是希腊人的命名法。我们管它叫做北斗七星。大家看北斗七星，从端点这边数第二颗星，中文名字叫开阳，端点那颗叫摇光。用肉眼就能看到开阳星的旁边还有一颗小星，中文名字叫“辅”，西洋名字叫80，就是大熊星座80（见图5-4）。古希腊对一个星座中的恒星这样命名，最亮的叫 α ，其次叫 β，根据恒星的亮度从大到小的顺序排过去。后来国际天文界沿用了古希腊的命名方式，希腊字母排完后，再用小写的拉丁字母排序a, b, c, d.....，接着再用大写的拉丁字母排A, B, C, D.....。此外国际天文界还有另一种命名法，把星座中的恒星按它们在天空中的方位（赤经、赤纬）编号排列，用数字1, 2, 3.....往下排，辅的西洋名字大熊座80，就是这样命名的。

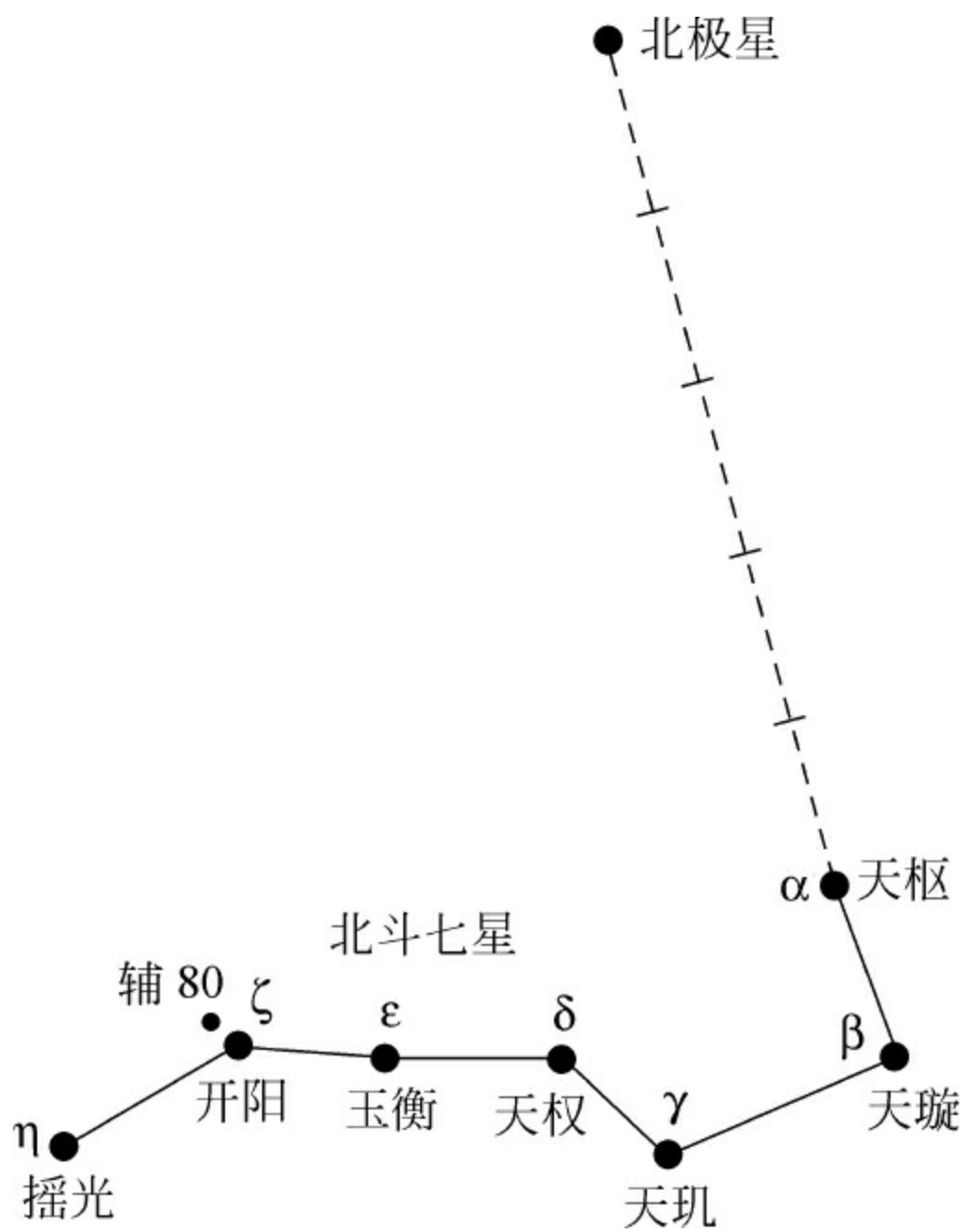


图5-3 北斗七星与北极星



图5-4 北斗七星中的开阳

多个太阳的太阳系

人们发现开阳和辅是一对双星。我们通常看到的双星有两种情况，一种情况是这两颗双星没有物理关联，它们只不过从地球上是在同一个方位上，其实它们两个前后距离差得很远，这种双星我们一般兴趣不大。我们感兴趣的是有物理关联的双星。观察发现辅和开阳是有关联的，这是一对真正的双星，围绕它们的质心转动，好像是由两个太阳组成的太阳系。有了望远镜以后发现开阳本身是双星，有两颗；辅也是双星，也有两颗，围绕共同的质心旋转。再仔细看，开阳星这对双星中的每一个又都是由两颗恒星组成的。所以这个太阳系一共有六个太阳（图5-5）。像这样的恒星系，肯定有行星，但是有高级生命的可能性恐怕不大。按照现在人类的观点来看，不大容易有高级生物。因为行星很可

能会在恒星当中穿来穿去，温度变化非常剧烈，说不定海洋都开锅了！高级生命可能忍受不了。但这是我们人类目前根据现有知识产生的看法。

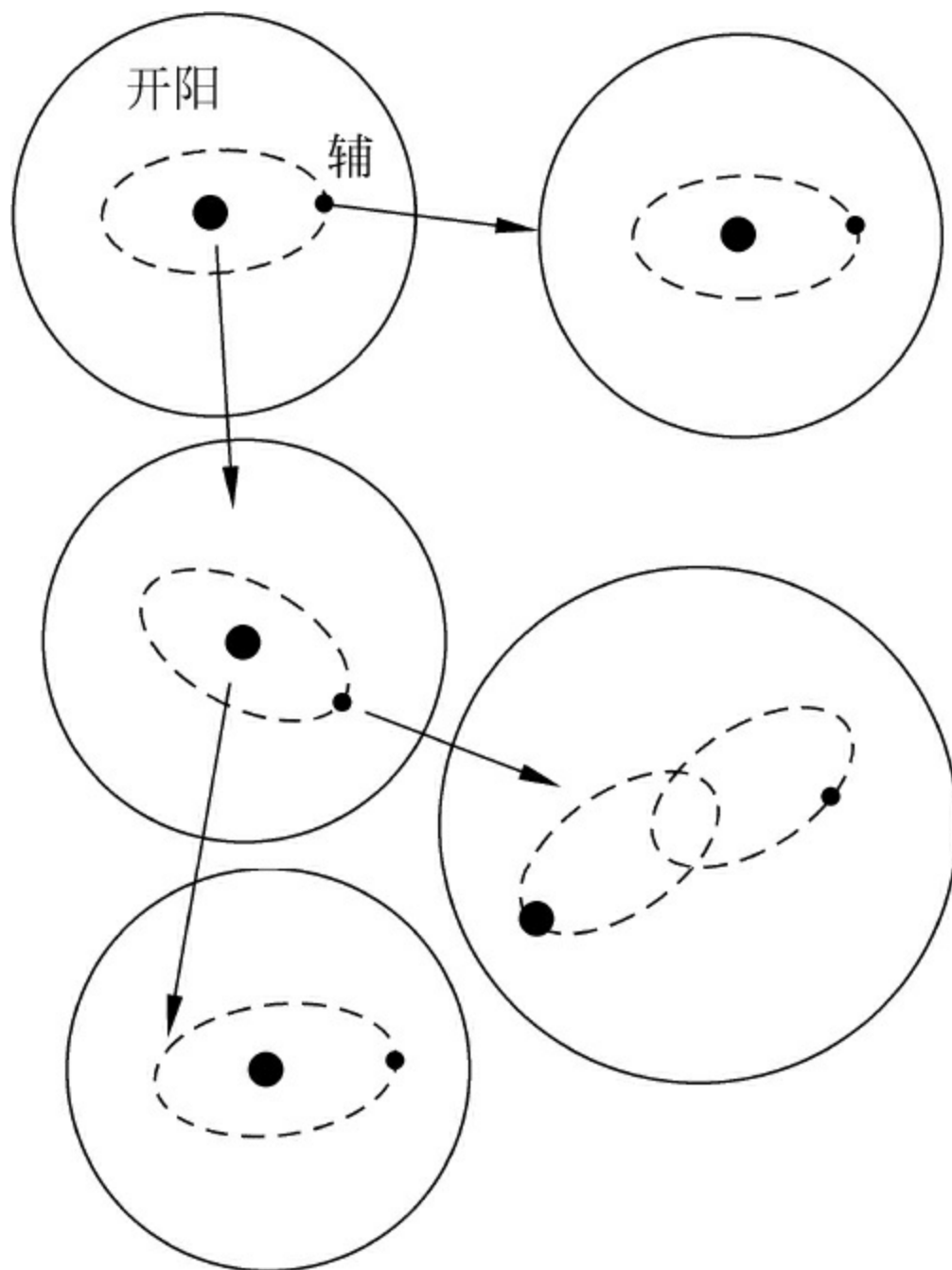


图5-5 开阳与辅

我们的银河系

我们都知道，太阳系有太阳，有行星，还有卫星，具有成团的结构。有的“太阳系”有几个太阳构成，有的“太阳系”只有一个太阳。这些“太阳系”还构成更大的结构，比如说有的构成星团，其中一种叫球状星团，如图5-6所示，由几万或者几十万颗恒星组成。还有一种疏散星团，是由几十颗到上千颗恒星组成。

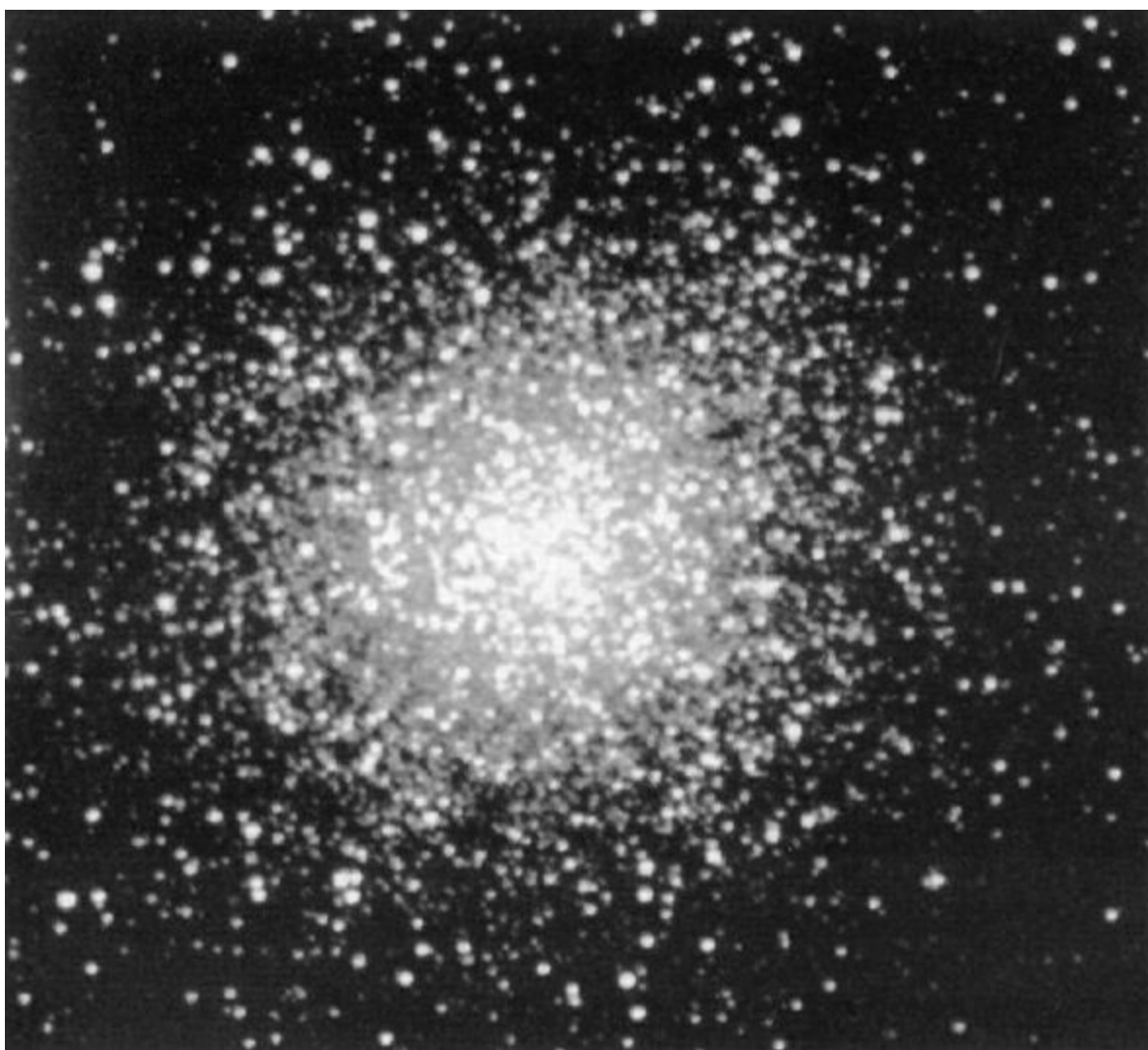


图5-6 球状星团

图5-7是银河系的侧视图，主体是银盘，大量的恒星聚在银盘上。不过球状星团不在银盘上，在银盘之外，由上万颗恒星组成。银盘上还有一些疏散星团，但是更多的是较为独立的恒星，组成一个一个的太阳

系，然后再一同组成银河系。因此我们看到在银河系以下的层次当中，好像物质结构都是成团的结构。图5-8是银河系的俯视图。

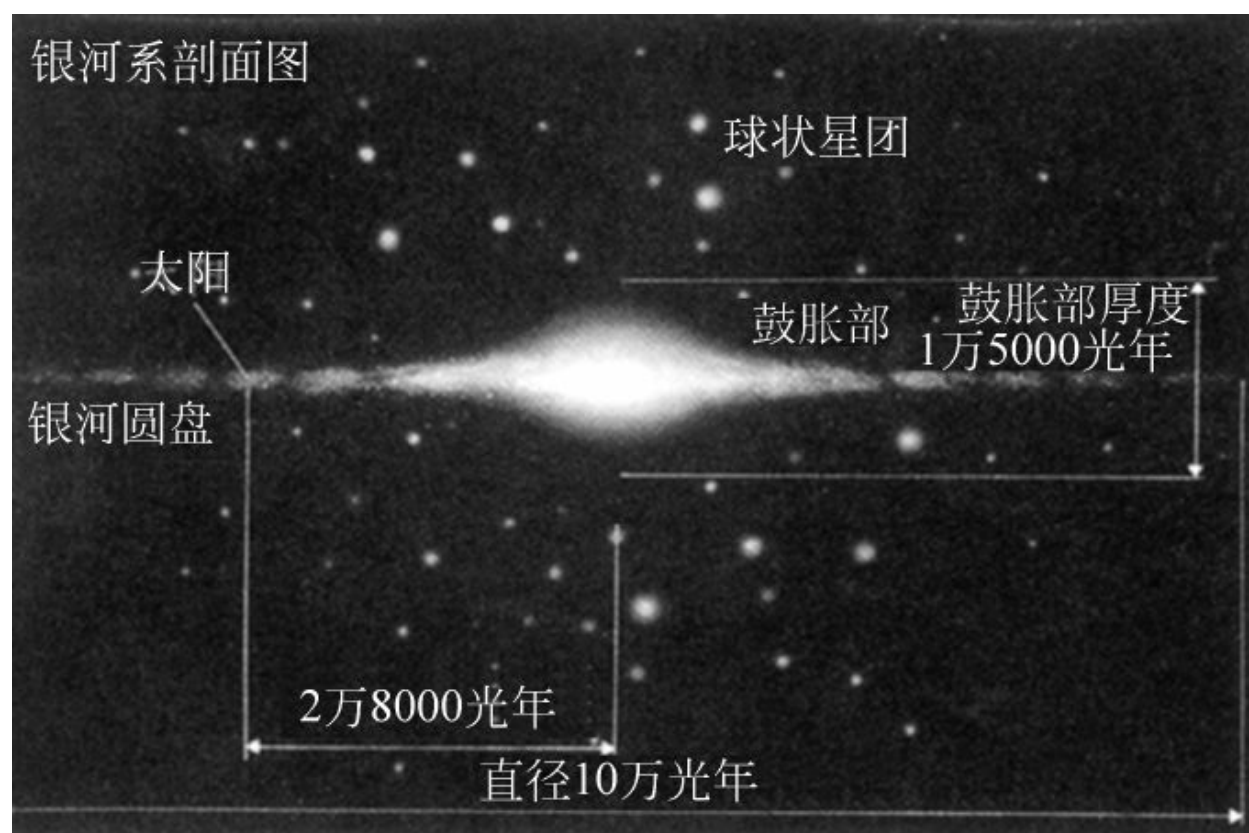


图5-7 银河系侧视图

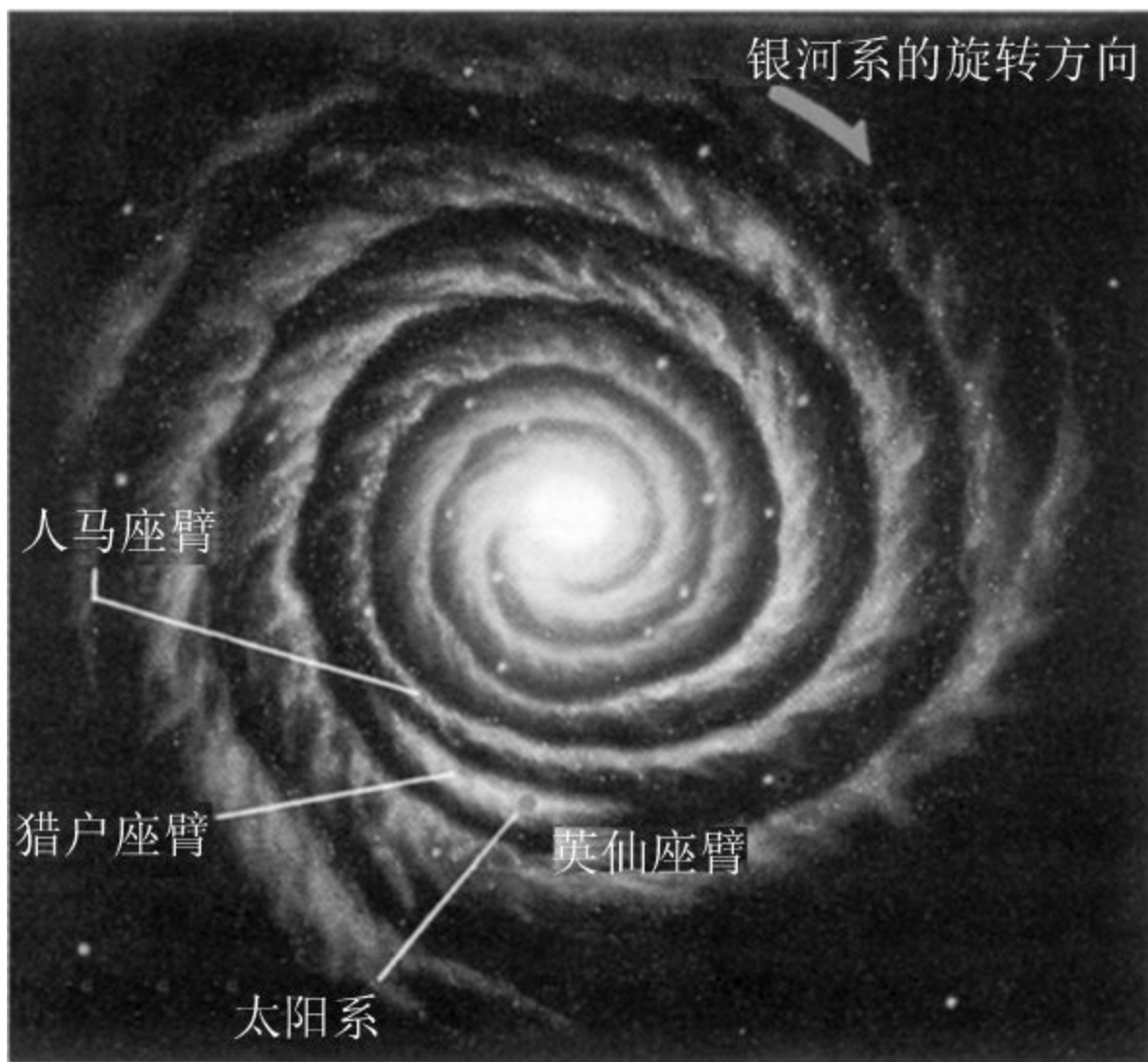


图5-8 银河系俯视图

遥远的河外星系

后来人类发现，在我们银河系之外还有其他的银河系，比如在仙女座那个方位有一个星系，肉眼就可以看到，好像一个模糊的光斑。这个星系称为仙女座星系，很大也很漂亮。它离我们比较近，有220万光年，直径是16万光年。我们银河系直径是10万光年，它比银河系略大一点。它所处的角度很好，人们拍的照片（图5-9）看得比较清楚，可以说非常漂亮、非常清楚。

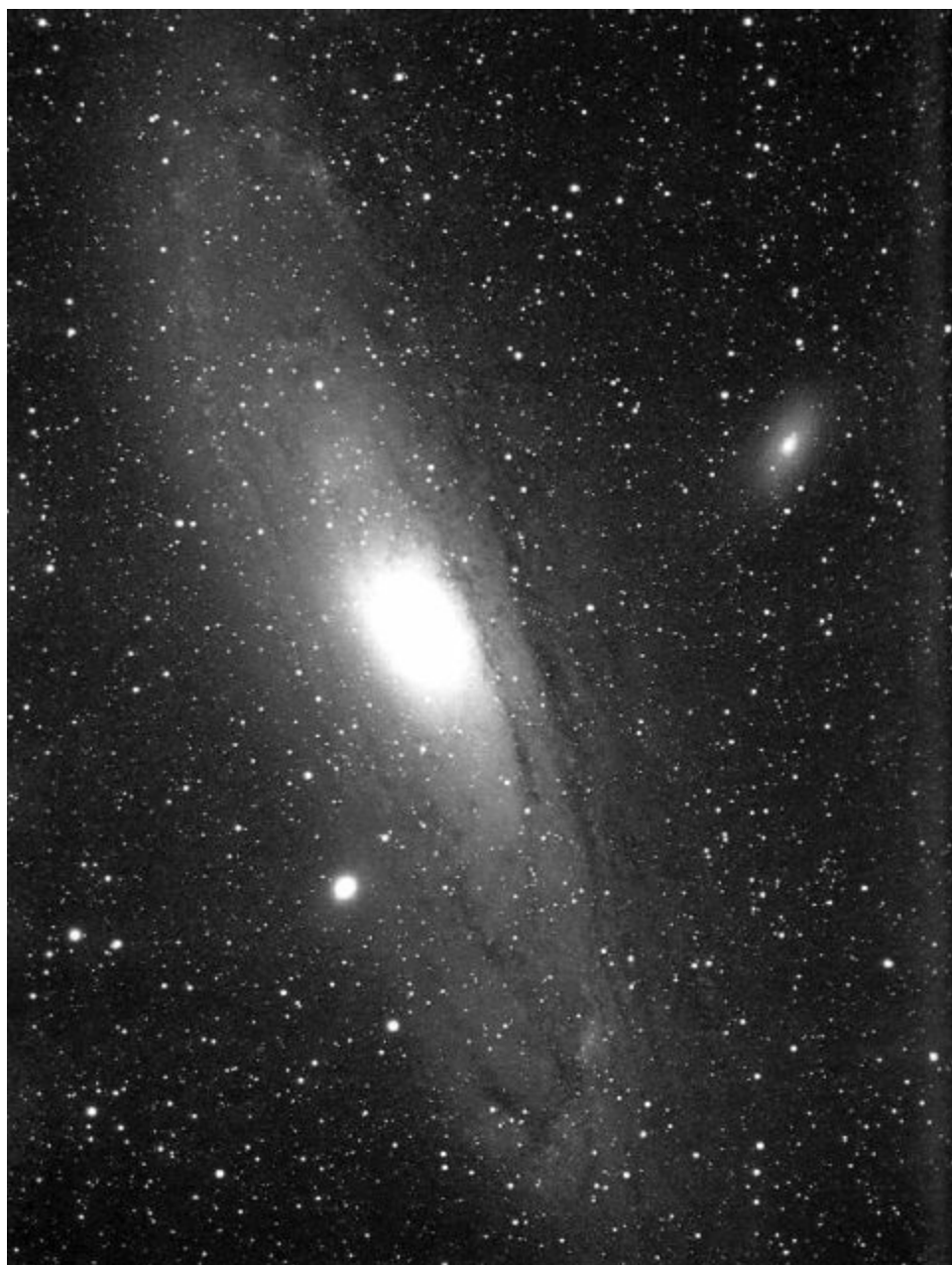


图5-9 仙女座星系

从地球上肉眼可见的还有麦哲伦云，大麦哲伦云和小麦哲伦云（图5-10）。这两个麦哲伦云，距离我们很近，比仙女座星系还近，它们分别距离我们16万光年和19万光年。为什么叫麦哲伦云呢？这两个星系在北半球看不见，麦哲伦环绕地球航行的时候，曾经在赤道以南航行，在赤道以南可以看见南天的星空，这些星空在北半球看不到。当麦哲伦的

船队穿越美洲最南端的海峡时，随船的天文学家在南半球的天空发现了这两个星系，就把这两个星系命名为大麦哲伦云和小麦哲伦云。



图5-10 大小麦哲伦云

星系群与星系团

大、小麦哲伦云、仙女座星系跟我们的银河系以及其他的几十个星系一起组成一个星系群，围绕它们共同的质心旋转。这就是说，像银河系这样的星系，它不但自身是成团的，而且还跟别的星系一起构成星系团。星系团如果只有几十个星系，就叫星系群。如果有更多的，比如说上千个星系的，就叫星系团了。可见星系也是成团结构。在星系团之上还有超星系团。图5-11是用望远镜看到的星系团和星系群，一个一个的星系，五颜六色，非常漂亮。星系团和超星系团直径基本上在 10^7 光年，也就是一千万光年这个数量级上，在这样的尺度上，物质还是成团结构的。比它再大， 10^8 光年，也就是一亿光年的尺度上，宇宙中一个一个个的超星系团都是均匀分布的，在空间是均匀分布的，而且近处看是

均匀分布的，往远处看也是均匀分布的，都是均匀各向同性地分布着。所以宇宙物质的成团结构大概是在小于一亿光年的尺度上，比如在一千万光年的尺度上是成团的，在那以上就不是成团的了，而是均匀各向同性地分布着。

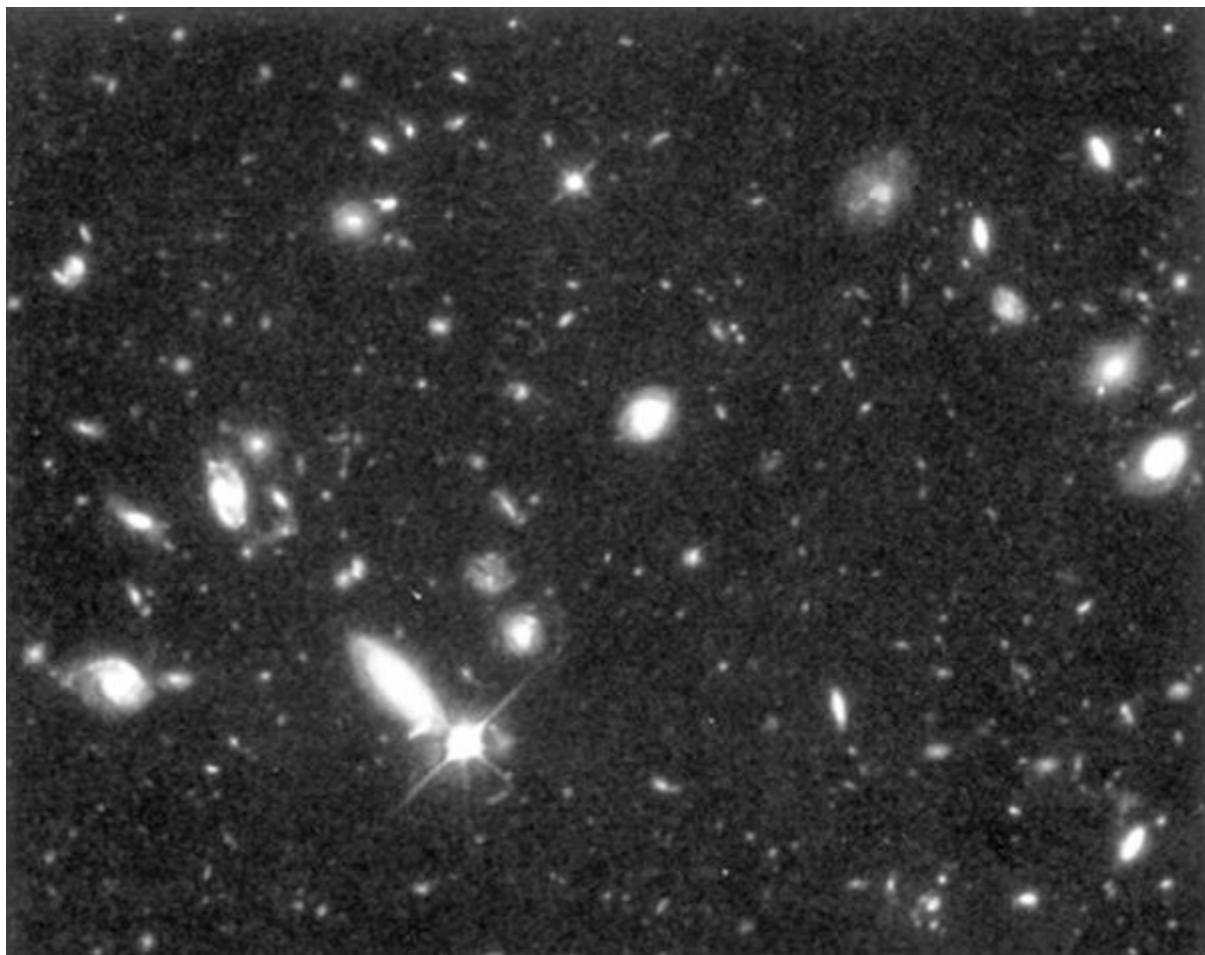


图5-11 星系团与星系群

宇宙物质的成团结构

我们对于宇宙的了解，大体上可以小结如下，太阳系的直径是一光年，它的半径是半光年。就是它的范围不仅伸展到海王星和冥王星处，太阳引力的控制范围要远远比那个遥远。像银河系这样的星系，每一个的直径大概都是十万光年，在 10^5 光年左右。星系团和超星系团一般在 10^7 光年这个量级，也就是一千万光年的量级。观测表明在更大的尺度

上，宇宙是均匀各向同性的。这就是爱因斯坦那个时代人们所知道的宇宙结构。

2. 爱因斯坦的静态宇宙

从时空弯曲认识宇宙

爱因斯坦为什么会去研究宇宙呢？他的广义相对论认为万有引力是时空弯曲的表现。广义相对论是1915年，也就是一战期间完成，1916年发表的。那时候，量子论正在蓬勃发展，很多人认为是不是可以把广义相对论用于量子论的研究，比如说对原子光谱会不会有修正。

现在我们知道有力有四种：原子核内部的强相互作用力，还有电磁力，弱作用力和万有引力，一共四种。但是那个时候只知道电磁力和万有引力。

量子力学里原子能级的跃迁，从一个能级跃迁到另外一个能级，会发射光谱线，这是电磁相互作用的结果。电子和原子核之间的电磁相互作用，若想用广义相对论来修正，就要考虑电子和原子核之间的万有引力的影响，也就是时空弯曲的影响。它能够造成多少修正呢？爱因斯坦考虑电磁相互作用的强度，比如说一个电子和一个质子，它们的电磁作用的强度和它们的万有引力的强度相比差 10^{37} 倍。所以他就觉得时空弯曲对量子力学的影响，对光谱线的影响，是根本看不出来的。因此把广义相对论应用于量子力学的想法实际上是根本不行的，完全没有用。

爱因斯坦觉得可以用广义相对论来研究什么呢？他觉得可以用来研究宇宙。为什么呢。因为组成宇宙中星系的物质都是由原子、分子，或者说中子、质子、电子构成的，虽然很多基本粒子带电，但是整个星系都是电中性的。所以电磁相互作用就被掩盖掉了，那么突出出来的就是万有引力。因此他就主张把他的广义相对论应用于对宇宙的研究。

均匀的宇宙

爱因斯坦首先根据我们前面所说的那些对宇宙的认识，提出了一个

宇宙学原理，说在宇观尺度上，物质的分布始终是均匀各向同性的。

宇观尺度就是 10^8 光年以上，就是比一亿光年更大的尺度。在这样的大尺度上，物质的分布始终是均匀各向同性的。也就是说你用望远镜看任何一个方向，物质密度都差不多，星系团的密度都差不多。所以是均匀各向同性的。

怎么知道“始终”是均匀各向同性的呢？因为你要观察离我们一光年的一颗星，我们看到的是它一年以前的样子。距离我们十光年的星，我们看到的是它十年前的样子。

所以望远镜不仅在看远方，而且也在看历史。越远的星系，我们看到的越是它古老的景象。由于你看到不管远处还是近处，星系团都是均匀各向同性分布的，所以爱因斯坦就提出来一个宇宙学原理，说宇宙当中物质的分布始终是均匀各向同性的。

静态的宇宙

这是爱因斯坦根据当时的天文观测提出来的一种对宇宙的看法。按照这个看法，我们的宇宙，过去均匀各向同性，现在也均匀各向同性。当时的望远镜还不能看得非常远，用望远镜看，这些星系团的情况都差不多。所以他就觉得，我们这个宇宙可能从总体上讲，没有大的变化。

虽然恒星和星系团都不断地成长、衰老，又不断有新的星系产生出来。老的灭亡，新的产生。但是总的密度不会有什么大的变化，物质密度不会有什么大的变化。所以他觉得我们这个宇宙应该是不随时间变化的，他就想从他的广义相对论方程式（5.1）把这个宇宙模型解出来。

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu} \quad (5.1)$$

有限而无边

大家看，式（5.1）这个方程的左端表示时空曲率，表示时空的弯曲情况，右端表示物质的能量动量的分布。爱因斯坦方程是十个二阶非线性偏微分方程组成的方程组，非常难解，谁要得出一个解，就可以用

他的名字命名。现在得出来的解还是有一些，但是真正有物理意义的并不多。

爱因斯坦想从这个方程解出他的静态宇宙模型。但是我们知道解微分方程除去有了方程之外，还得知道初始条件和边界条件，就是你研究的这个场，初始时候是什么情况，边界是什么情况。

比如说我们要研究这个屋子里的电磁场，我们有了麦克斯韦方程组，是不是就能解决屋子里电磁场的分布呢？不行！你还得知道边界条件，就是我们的墙壁是什么材料组成的，它是金属还是非金属，这叫边界条件。另外还得知道初始条件，比如初始时刻屋子里边的电磁场是怎么分布的。如果你知道了初始条件又知道了边界条件，就可以用电磁场方程把以后每一时刻这个屋子里的电磁场的分布全都推定出来。

所以对于爱因斯坦来说，还需要有宇宙的初始条件和边界条件。不过他这个模型是静态的，不随时间变化，所以初始条件就可以不要了，或者说初始条件就是现在这个样子。那么边界条件呢？宇宙的边是什么样啊，这个事情很难回答，假如有人说边界是什么样，可能有人就会问，那么边界外算不算宇宙呢？这也是很麻烦的事。

爱因斯坦倒是想得比较简单，他说我这个静态宇宙是有限无边的，没有边，所以也就不需要边界条件了。大家会想，有限不就是有边吗？像这个桌子，面积有限，就是长乘宽。用手一摸就是边，有限有边。大家都知道欧几里得平面无限无边。一个有限有边，一个无限无边，怎么还会存在有限无边的情况呢？

爱因斯坦说，你们看一个篮球的表面。这个篮球的表面面积是 $4\pi r^2$ ，是有限的。一个二维的生物在上面爬来爬去，永远爬不到边，这就是一个二维的有限无边的空间。爱因斯坦建议大家充分地发挥想象力，想象我们的三维空间是有限无边的。他认为三维空间应该是一个超球面。超球面可不是个实心球啊，那是四维时空中的一个三维的球面。我们的宇宙就是这样的，四维时空其中的一个三维超球面。他认为时间

不停地走着，而这个超球面没有变化。

神秘的宇宙项

爱因斯坦想从他的方程推出这个模型来，但是努力了相当长时间都没有推出来。爱因斯坦这人确实很聪明，他很快明白了：时空弯曲程度很低的话，他的广义相对论就可以回到牛顿的万有引力定律。

所以这个方程实际上是万有引力定律的发展和推广，里面只有吸引效应没有排斥效应。一个系统如果只有吸引没有排斥的话，不可能稳定。所以爱因斯坦觉得用原来的方程不行。他就在原来的方程里加了一项，这项是常数 Λ 乘上一个 $g_{\mu\nu}$ 。 $g_{\mu\nu}$ 是什么啊？是度规，是与度量时空尺度有关的函数。新加的这一项叫做宇宙项。加进了宇宙项后，他得到了不随时间变化的有限无边的宇宙模型。这一项引进了排斥效应。

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu} \quad (5.2)$$

其实爱因斯坦早就知道这样的项是会引进排斥效应的。因为他在寻找广义相对论方程的时候，对方程的右端应该是 $T_{\mu\nu}$ 他觉得是没有问题的。但是左端是什么样的呢？他尝试了很长时间。起先格罗斯曼跟他合作，后来希尔伯特跟他合作，都是在寻找方程左边的函数形式。最终找到了 $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ 这样一项，找到了这个正确的结果。但是他也曾经试过 $\Lambda g_{\mu\nu}$ 这种形式，把它搁在左边。只把 $\Lambda g_{\mu\nu}$ 搁在左边的时候，他得不出与观测相符（例如水星进动）的时空弯曲情况。因为这一项只产生排斥效应，违背万有引力定律，所以他抛弃了这样的项。现在需要排斥项了，他又把这一项加进去。这个 Λ 叫做宇宙学常数，这个 Λ 通常是用大写字母表示。把这项加进来以后，引进了排斥效应，就得到了他所期待的静态宇宙模型。

当时的舆论界都轰动了，说我们伟大的爱因斯坦继狭义相对论和广义相对论之后，又把宇宙问题解决了。我们的宇宙是什么样子呢？是个

有限无边的东西。我相信当时所有的人几乎都搞不清爱因斯坦在说什么，搞不清他所说的这个有限无边的宇宙究竟是怎么回事。

3. 膨胀或脉动的宇宙

弗利德曼的突破

爱因斯坦认为他又解决了一个重大问题。但是过了不久，有个杂志社转给他一篇文章，是他没有听说过的一位苏联数学家弗利德曼写的。这个人用爱因斯坦没有宇宙项的方程，也就是原先的那个方程式

(5.1)，得到了一个膨胀的解，或者说一个脉动的解，脉动的解就是一胀一缩的解。弗利德曼主要得到的是脉动解，三维空间也是有限无边的。但是爱因斯坦的有限无边宇宙是不动的，他这个有限无边的宇宙是一胀一缩的。

爱因斯坦看了这个稿子以后，觉得这个工作是错误的。他对杂志社表示意见说，这个解是不对的。杂志社就不想发表，把审稿人意见转告了弗利德曼。审稿是背靠背的，弗利德曼并不知道审稿人是爱因斯坦。后来弗利德曼偶然从一个渠道听说审稿人是爱因斯坦，就给他写了一封信，解释自己的模型。爱因斯坦没有回信，看来爱因斯坦仍然坚持认为他的文章有问题。

弗利德曼没有办法，只好把稿件寄给了德国的一个小的数学杂志，弗利德曼本人是个数学家，他的解在数学上肯定没有问题。那个杂志就将文章登出来了。不过，由于那家杂志影响不大，他的文章登出来以后，没有引起大家的注意。

过了不久，在比利时有个叫勒梅特的神父，用带 Λ 的爱因斯坦方程，就是带有宇宙项的方程(5.2)，也求出了脉动或膨胀的宇宙模型，也得到了动态模型，这篇文章在另外的西方杂志上登出来了。文章登出来不久，哈勃定律就发现了。

哈勃的发现

哈勃定律是怎么回事呢？天文学家早就发现，宇宙中星系的颜色都有点变化，有的有点发红，有的有点发蓝。绝大部分都发红，有很少量的发蓝。那是怎么回事呢？大家想这可能是多普勒效应，说明这些星系相对我们有运动。

比如说一列火车朝咱们开过来了，声源朝我们运动，我们会觉得那声音很尖。一旦它开过去远离我们的时候，它的声音马上就钝下来了。这就是声学中的多普勒效应。光学也有同样的多普勒效应，当一个光源朝我们运动过来的时候，它的光的波长会变小，就是频率增高，光谱线会向蓝端移动，就是说它发蓝。如果远离我们就发红。

天文学家看到的绝大部分星系发红，只有极少量的发蓝。后来发现那些发蓝的星系都是跟我们银河系处在同一个星系群里边的。它们围绕星系群的质心运动，作相对运动，有的朝你运动，有的远离你运动，那确实是多普勒效应。但是我们的星系群以外的星系团、星系群，都是在远离我们的，光都发红。哈勃根据这些现象总结出来了一个公式，所谓哈勃定律

$$V = HD \quad (5.3)$$

D就是这个星系离我们的距离，V是它逃离我们的速度，H是一个比例常数，这个常数就叫哈勃常数，是哈勃的姓的第一个字母。

这个式子很简单，最初哈勃从观测的角度得到了距离跟红移的关系，如果把红移看成多普勒效应的话，红移量和V的关系很容易算出来，最后就可以得到 $V=HD$ 这个公式。这个公式支持了膨胀宇宙模型，支持了宇宙在膨胀。它表明远方的星系都在远离我们，而且离我们越远的星系逃离得越快。

爱因斯坦放弃宇宙项

哈勃定律支持了膨胀宇宙模型，所以爱因斯坦也觉得，看来膨胀宇宙模型是对的。爱因斯坦后来表了个态，宣布放弃自己的静态宇宙模

型，说他们的膨胀模型是对的。又说我的这个方程，正确的广义相对论方程，应该是原来那个没有宇宙项的，而那个有宇宙项的方程是错的。我不应该加进宇宙项，看来宇宙项不属于我的广义相对论，请大家以后就不要用了，把它忘掉吧。

可是他说让大家都忘掉，那大家能忘掉吗？很多人觉得还是可以有这么个东西啊，有相当一部分人不愿意放弃。而且研究这个带宇宙项的方程还可以求出一些新的解来。结果许多人还继续使用带宇宙项的方程，继续发表论文。

爱因斯坦很遗憾，说引进宇宙项看来是自己一生中所犯的最大的错误。这一情况就像“天方夜谭”里的那个渔夫，钓起一个魔瓶，一开盖魔鬼就出来了，出来了想塞回去也塞不回去了。

于是宇宙项就存在于相对论的理论当中了。此后研究相对论的人都是既研究不带宇宙项的方程，又研究带宇宙项的方程，两种方程都研究，至今发表的论文当中两种情况都有讨论。

4. 爆炸和演化的宇宙

神父的“宇宙蛋”

我们还要提一提这位勒梅特神父。他主张宇宙是膨胀的，当时就有人问了，你这个宇宙不断地演化，这跟上帝创造宇宙是不是有点矛盾啊？这个膨胀是上帝指挥的吗？这件事情好像还不太清楚。勒梅特解释说，其实很清楚，上帝当年创造的不是我们现在的宇宙，而是一个“宇宙蛋”，大概像乒乓球那么大。然后这个很热的蛋就膨胀起来，逐渐膨胀、散开、降温，演化成了我们今天的宇宙。他还用热力学对宇宙的膨胀进行了描述。这是一个非常重要的思想，就是用演化的观点来看待宇宙。

α 、 β 、 γ 的原始火球

这个时候有一个叫伽莫夫的物理学家亮相了。伽莫夫是苏联的大学

生，他和朗道一起被苏联派出国进修。这两个人确实都很了不起，朗道后来回国了，伽莫夫留在了西方，他研究核物理。

伽莫夫考虑，勒梅特讲的这个原始的、热的宇宙蛋是不是一个核火球啊？应该把核物理用进去。于是他就把核物理用到宇宙演化的研究中了。根据他的观点，最初宇宙是一个核火球，这个核火球逐渐膨胀开来，慢慢地降温形成我们今天的宇宙，这就是所谓火球模型。

伽莫夫指导他的研究生阿尔法进行研究，这个人的名字有点像希腊字母 α 的读音。伽莫夫自己的名字很像希腊字母 γ （伽马）的读音。伽莫夫这个人很爱开玩笑，正好他们研究所里有个叫贝塔的物理学家，于是他把贝塔也拉进来，阿尔法（ α ），贝塔（ β ），伽马（ γ ）三个人写了火球模型的文章。但主要贡献是阿尔法和伽莫夫的，特别是伽莫夫本人，思想主要是他提出来的。

但是这个模型遭到主张稳恒态宇宙模型的霍伊尔（就是后来被霍金挑错的那个人）讽刺，他说这个模型简直就是一场大爆炸，还不如叫大爆炸模型算了。结果大爆炸模型这个名称一直沿用至今。现在大家叫火球模型的反而少了。

伽莫夫和勒梅特两个人都说宇宙在演化。大家知道，生物学刚开始的时候只讲植物、动物分类，到了达尔文的时候诞生了进化论，对人的研究后来也出现了进化论。历史研究表明：人类社会和人类文明也在进化。勒梅特和伽莫夫现在告诉我们：宇宙同样是发展演化的，从静态到了动态，这是人类思想上的一次大飞跃，对宇宙认识的大飞跃。

勇敢加天才

现在我们回过头来再看一下哈勃定律。图5-12是哈勃最早给出的那张图，得到哈勃定律的图。纵坐标是远方的星系逃离我们的速度，每秒多少公里。横坐标是这些星系离我们的距离，这里“秒差距”是一个天文学单位，搞天文学的人都用秒差距，他们觉得比较方便、正规，但是这个东西一般人觉得不太直观。大家对光年更熟悉，一个“秒差距”大概

是三个多光年的样子。你看图上的观测点多散啊，哈勃很勇敢地“噌”就一条直线画过去了，成正比。后来有些人想，哈勃怎么回事啊。有的人猜：他是不是知道膨胀宇宙模型啊？如果知道膨胀宇宙模型的话，成正比的直线，正好跟膨胀宇宙模型一致。但是也有另外一种可能，就是他抓住了主要矛盾。因为他是搞观测的，他知道这些数据的误差到底有多少，他的观测点当中到底有多少是非常可靠的，可靠到什么程度，所以他抓住主要东西勾勒出了这条直线。如果你详细描绘的话，这条线将是复杂的曲线，根本找不出什么规律来，乱七八糟的。哈勃则把主要的实质描出来了。事实上随着实验观测越来越精确，我们发现新的观测点也越来越靠近这条线。

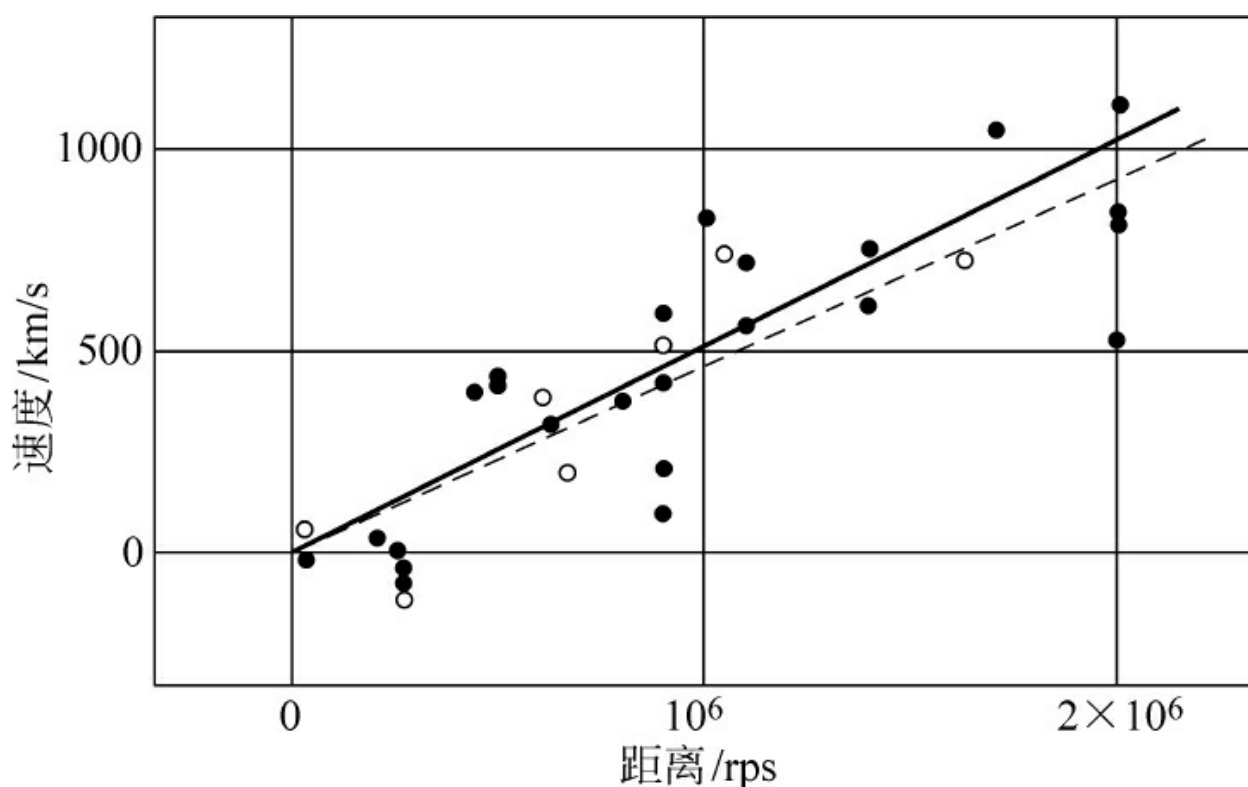


图5-12 最早的哈勃图

搞实验研究的人还真的要注意，一方面要关注理论，但是千万不能像有的人那样，理论上是什么你就造出什么来，那是不行的。但是你在

做实验的时候也要注意抓主要矛盾，抓住主要规律。

氢丰度的支持

这个火球模型提出来的时候，伽莫夫就说，我的火球模型是有观测支持的，第一是哈勃定律，它表明宇宙一直在膨胀。大爆炸以后的膨胀确实应该是这个样子。另外一个就是氢丰度，就是说宇宙中氢元素的含量，按他的火球模型应该是20%多。

宇宙刚开始生成的时候，主要的物质形态是以氢为主的气体。因为那时候的温度高，氢就聚合成氦。然后再继续膨胀，温度就降下来了，氢聚合成氦的热核反应，也就基本停下来了。

根据伽莫夫的计算，这个时候的气体应该含有20%多的氦，70%多的氢，这就叫氢丰度。根据他的理论计算出来的这个值，是跟观测值大体相同的。当然，后来这些气体聚集成团以后，会往里收缩，形成一颗颗恒星。收缩过程当中大量的万有引力势能转化成热能，温度重新急剧上升，升高到上千万度、上亿度之后，重新点燃气团中心的热核反应，氢再烧成氦，就像我们的太阳这样发出光和热。但是这样生成的氦，与在宇宙早期生成的氦相比微乎其微，很少，所以氢丰度是对火球模型的一个支持。

大爆炸的余热

另外他还说，我们的宇宙既然原来是一个原始的大火球，那么今天的宇宙不可能是热力学温度零度，一定还有大爆炸的余热保留下来。他估计这个余热大概是热力学温度五度左右，有人猜测是热力学温度十度左右。

这个大爆炸余热长时间没有找到。伽莫夫是在1948年提出火球模型的，十几年后，1964年才发现了大爆炸的余热。当时有一些搞相对论的专家想找大爆炸的余热，用各种办法寻找，看宇宙空间中有没有余热，但没找着。这时候另外两位射电天文学家却在无意中找到了。

射电天文学家是干什么的呢？他们研究来自宇宙空间的无线电波。

我们知道，人的肉眼看到的都是可见光，后来有了望远镜也是看星星射来的可见光。但是有些天体不发可见光，只发射无线电波，还有一些天体既发可见光，又发无线电波。所以来自宇宙空间的无线电波，对于我们也很重要，我们也能够从中知道宇宙空间的很多信息。

当时有两位美国射电天文学家，一个叫彭齐亚斯，另一个叫威尔逊。他们设计了一套天线装置，打算寻找来自宇宙空间的无线电波。他们设计好这套装置以后，尽量想办法降低噪声，提高装置的灵敏度，先把它调好，再进行观测。但他们发现调到一定程度，噪声就降不下来了。他们想肯定是天线有问题，于是把整个天线拆开，进行检查。发现在天线的核心部位，鸽子修了一个窝，还拉了一堆鸽子粪。论文当中写了他们清洗鸽子粪，把窝也拆掉了。当然讲述这个过程的时候，他们采用了很文雅的词来描述，他们说发现了一堆鸽子的白色分泌物，然后把这些分泌物给洗掉了。但是进行了这些操作之后，噪声依然存在。

这回他们明白了，这个噪声可能是宇宙空间本来就有的，并不是他们的设备造成的。这个噪声是热力学温度 2.7K 的微波背景辐射。我们知道温度很高的时候，物体主要的辐射处在可见光波段，温度低下来，可能就跑到波长较长的红外波段去了，再低就到微波波段了。热力学温度 2.7K （我们通常说 3K ）的热辐射，就是微波波段的辐射。微波背景辐射发现之后，大爆炸模型得到了公认。在20世纪60年代后期和70年代，大爆炸宇宙学迅猛发展起来，大家普遍相信了这个模型的正确性。

膨胀没有中心

我们经常有一个问题，就是从地球上看到，宇宙空间在向外膨胀，所有的星系都在远离我们，是否我们这里就是爆炸的中心呢。对不对？不对。实际上，你站在任何一个星系上都会看到别的星系在远离自己。为什么呢？请大家看图5-13中这个吹气球的小孩，这张图就是用来解释这个问题的。你看这个气球，气球上有很多墨水点，这些墨水点表示一些二维的星系，这个气球就是一个二维空间。小孩一吹，这个二维

空间就膨胀了。你看，对于任何一个墨水点来说，其他的墨水点都在远离它，对不对？所以在这个膨胀的宇宙中，没有膨胀的中心，或者说每一点都是膨胀的中心。

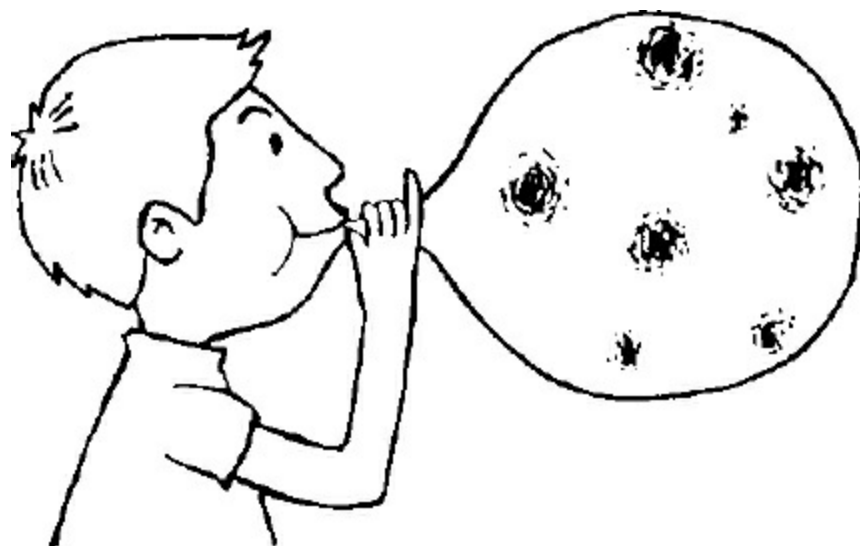


图5-13 膨胀的宇宙示意图

宇宙的创生、暴胀与演化

图5-14讲解了宇宙演化的过程，对宇宙演化的描写较详尽。总的说来就是，宇宙最初的时候是无中生有的，时空和物质一起从虚无当中冒出来。需要说明的是，宇宙刚刚从虚无中诞生出来的那一段时间（ 10^{-43} 秒以内），既分不清上下、前后、左右，也分不清时间和空间，只是在 10^{-43} 秒之后这些东西才能够分清。

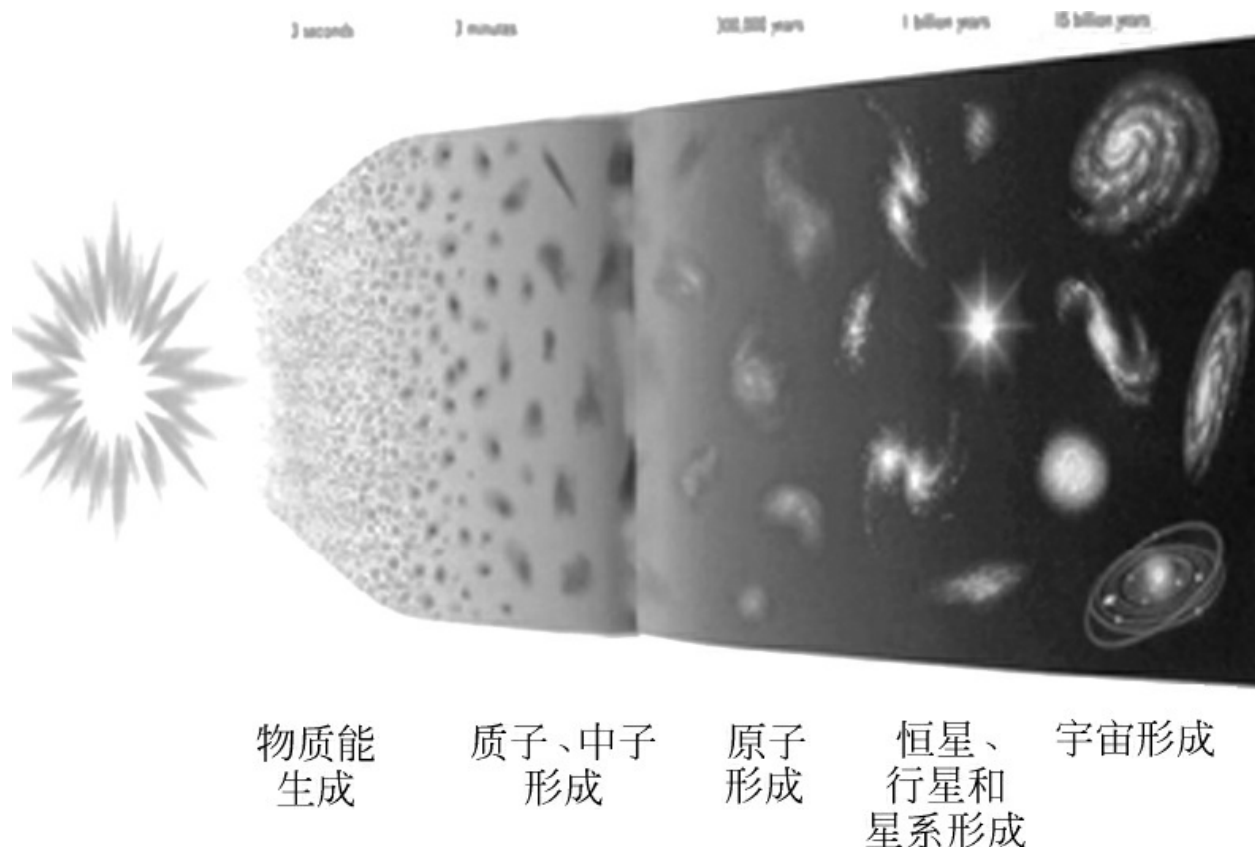


图5-14 宇宙演化的示意图

接着先经历了一段比较平稳的膨胀、降温过程，然后宇宙出现一个以真空能为主的时期，这时宇宙进入暴涨阶段，一个迅猛膨胀的阶段。在这一演化过程当中，原有的真空演变为“过冷的”、不稳定的假真空状态。接着假真空态会一下子跃迁到能级较低的真的真空态，此时大量的真空能转化成为物质能，从“虚无”中涌现出来，使宇宙回升到高温状态，然后便又恢复为平稳膨胀的状态，在膨胀中逐渐降低温度。温度逐渐降低以后，出现了夸克、胶子和轻子；然后就是质子、中子的形成；然后是原子核的形成，原子、分子的形成；更后来就是恒星、行星和星系的形成、演化；生命的出现，生物的进化，有些生物还发展成为比较高级的生物，产生了思想和意识，其中一些高级生物还在那儿提问题，问：我们为什么会从宇宙中产生呢？比如我们人类，就是这样的生物。

宇宙膨胀还是脉动

现在我们知道宇宙有两种演化方式。如果我们的三维空间是负曲率的，曲率 $k < 0$ ，是一个伪超球面，或者是平直的，曲率 $k = 0$ ，是一张超平面，那么我们宇宙的三维空间就是无限无边的，它会永远膨胀下去。还有一种情况是，我们的三维空间曲率是正的， $k > 0$ ，那么它是一个超球面，这个时候它的空间是有限无边的，而且是脉动的，不断膨胀又收缩。你看图5-15，这个图就画出了两种情况：在 $k < 0$ 、 $k = 0$ 的时候它是永远膨胀下去的，如果 $k > 0$ ，就是三维空间的曲率大于0，那么空间就膨胀然后又收缩。收缩时温度当然又重新上升。如果是永远膨胀的话，就会越来越冷，逐渐趋于热力学温度零度，却又达不到热力学温度零度，它就这么演化下去。如果是一胀一缩的话，就有一个空间不断膨胀、温度不断降低的时期，然后空间又重新收缩，温度又重新高起来。

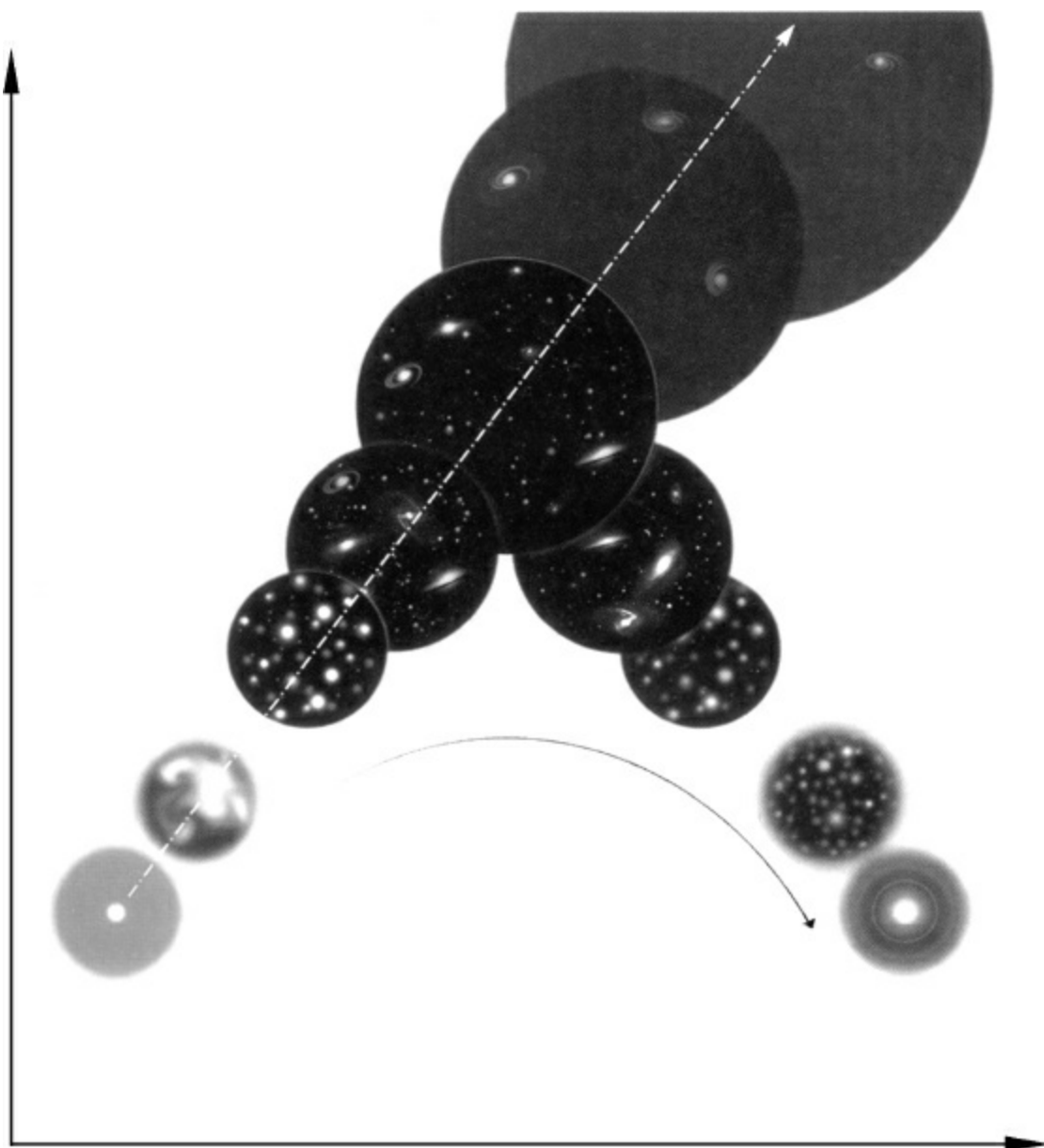


图5-15 宇宙膨胀的几种情况

那么我们的宇宙到底将来会怎么样呢？研究表明，我们宇宙中的物质有一个临界密度。这个临界密度是每立方米三个核子，或者说三个氢原子。如果宇宙中的物质平均密度大于这个临界密度的话，这个宇宙膨胀到一定程度就会收缩，宇宙将是脉动的。因为物质密度比较大的话，物质间的万有引力作用就比较大，也就是说，吸引效应会比较大，大到一定程度，就能够使宇宙膨胀转变为收缩。如果物质密度达不到这个临

界密度，那么宇宙就会永远膨胀下去。虽然吸引效应能促使膨胀减速，却不能使膨胀速度降到零，更不能使膨胀转化为收缩。

最初得到的观测数据表明，宇宙中物质的密度远小于临界密度，所以认为宇宙会永远膨胀下去。但是，另一些人研究河外星系的红移，发现这些星系虽然在远离我们，但是远离我们的速度下降得很快，即所谓的减速因子很大，这又似乎表明宇宙的膨胀将逐渐停止并转化为收缩。这就是说，研究减速因子得出的结论与研究临界密度得到的结论相反。那个时候有很多研究讨论纠缠在这个问题上。也就是说大家不太清楚，我们宇宙中的物质密度到底是多大，究竟是大于临界密度还是小于临界密度。有人猜测宇宙中物质密度可能正好等于临界密度，但又觉得有些怀疑。

5. 宇宙居然在加速膨胀！

近年来上述矛盾更加大了。大在哪儿呢，就是通过对一类超新星离我们距离的观测，发现宇宙目前根本不是在减速膨胀，而是在加速膨胀，从大约60亿年前开始，宇宙就从减速膨胀转变为加速膨胀了。这一发现完全颠覆了宇宙膨胀一直是逐渐减速的认识。最近颁发的诺贝尔奖就与这一发现有关。

天文学上的测距：量天尺

我们知道，天文学上测距有很多种方法。最古的时候是采用三角法，比如测月亮到地球的距离怎么测，我们从地球上的两点去观测月亮，形成一个三角形，地球上的基线长度是知道的，然后再测定两个观测点到月亮的连线与基线的这两个夹角，就可以算出月亮离我们的距离。后来还有其他很多种方法。比较著名的有一个变星方法。有一种变星叫造父变星，这种变星的变光周期跟它的真实亮度有关系，我们知道它的变光周期，就可以知道它的真实亮度，然后再用眼睛看它的视亮度，视亮度和真实亮度有差别的原因，就是它距离我们远。依据造父变

星视亮度和真实亮度的差值，我们就可以判定这个造父变星所在的星系离我们的距离了。所以造父变星就成为了一种标准烛光，一种量天尺。

I a型超新星：新的标准烛光

现在的问题是什么呢？天文学家突然发现又有一个东西可以作为测量星系距离的标准烛光。这就是 I a型超新星爆发。不同超新星爆发的规模是很不一样的，你不能随便选一颗超新星用它爆发的亮度作为标准烛光去判定它离我们的远近。因为超新星的个头可以差异很大，它炸完以后有没有剩下星体，剩下什么星体，剩下一颗中子星，还是剩下一个黑洞，这些剩余星体有多大，都是不一定的。所以它爆炸的规模很不一样。但有一种超新星爆炸的规模，许多天文学家认为是一样的。什么呢？就是 I a型超新星。

这种恒星本来它不会超新星爆发，它形成的是一颗白矮星。我们知道形成白矮星以后就不会爆发，然后它就会冷却下来，慢慢地变成黑矮星。但是如果这个“太阳系”，白矮星所在的“太阳系”，里头还有其他恒星，也就是说它处于一个双星或聚星（多个恒星组成）系统，那么其他恒星的物质就会被白矮星吸引过来围着它转，很多物质被吸积进去，于是白矮星质量逐渐增加，但白矮星有个质量上限，就是钱德拉塞卡极限——1.4个太阳质量。超过1.4个太阳质量的白矮星是不稳定的，它内部的泡利斥力抵抗不住万有引力，就会塌下去而爆炸，形成超新星，这就是所谓 I a型超新星。它是白矮星吸积了大量物质以后，质量超过了钱德拉塞卡极限，也就是1.4个太阳质量的时候，形成的爆炸。

这样爆炸形成的超新星，第一，它的质量不足以形成黑洞，也不足以形成中子星。结果是什么呢，当然是全部炸飞。它没剩下渣滓，全部炸飞掉。第二，个头都差不多，都刚刚超过1.4个太阳质量就炸了。所以许多人认为，这类超新星爆发的规模和亮度都差不多，是可以作为标准烛光的。

天文学家利用它作为标准烛光，重新测定遥远星系离我们的距离，

并和这些星系的逃离速度比较，就发现宇宙在诞生的初期，膨胀确实是逐渐减速的。但是在60亿年前突然变成加速了，从减速膨胀变成了加速膨胀。

排斥效应：透明的暗能量

怎么来解释这种加速膨胀呢？加速膨胀表明必定有个排斥力在起作用，于是有人就提出一个观点，说宇宙中可能存在暗能量。所谓暗能量是这样一种物质，它的压强是负的，所以它起了推动膨胀的作用。而且，它在宇宙中是均匀分布的，在宇宙膨胀过程中，暗能量的密度是保持不变的。但是随着宇宙空间体积的不断胀大，宇宙中暗能量的总数不断增加，它所产生的排斥效应就越来越强，终于压倒了普通物质间的万有引力造成的吸引效应，使宇宙膨胀从“减速”转变为“加速”。而且，这种暗能量还有一个奇特性质：不参加电磁相互作用。不参加电磁相互作用就使你看不见它，它既不发光，也不挡光，它对光是透明的。这种物质就叫暗能量。

吸引效应：透明的暗物质

在此之前，天文学家还猜测有一种暗物质，暗物质是什么呢？它产生吸引作用，它产生万有引力的吸引效应，但是也不参加电磁作用，对光也是透明的，看不见。暗物质的最初提出大概是在研究银河系这类星系转动的时候。

用牛顿力学来研究银河系中恒星的转动就会发现问题。银河系的银盘的旋转角速度应该与向心加速度有关。它应该受到一个向心力，这个向心力就是银河物质产生的万有引力。但是从银河系外围的这些恒星的转动速度来看，作为引力源的银河物质似乎并不都集中于银心，而是呈晕状分布，并且它们产生的万有引力，应该比我们看见的银河系中心的发光物质所能产生的引力要大很多。但是我们又只能看到那么多发光物质，而且这些发光物质比较集中于银心，不呈晕状分布。

有人说是不是尘埃和气体的质量没有考虑进去呢？不是。因为尘埃

和气体都会挡光，光过来了有些气体不仅会被照亮，而且自身也会被激发而发光。这是因为恒星的光照到它们上面以后，把它们的分子能级激发了，所以它们本身也发光，我们能看见。即使一些尘埃干脆把光挡住了，我们也能看见有黑不溜秋的一块。但是现在这种晕状分布的物质是你看不见的，它不参加电磁作用，跟玻璃一样，跟以太一样，透明的，光就穿过来了。但是它有万有引力，这种东西就是所谓暗物质。在研究远处星体的引力透镜和其他效应时，天文学家也产生过存在暗物质的推测。

当代的两朵乌云：暗物质与暗能量

暗能量跟暗物质都是透明的，暗物质是成团结构的，基本上是聚集在有恒星、有星系的地方，比如银河系的中心附近，许多人推测存在大量的、我们看不见的这种暗物质。暗物质是跟普通物质聚在一起的，是成团结构的。而暗能量呢？是在宇宙当中均匀分布的。在宇宙膨胀过程中，它不是密度减小，而是密度保持不变。所以随着宇宙的膨胀，宇宙中的暗能量总量就越来越多，排斥效应就加大了，因此宇宙就变成了加速膨胀。

对这两个问题，很多人进行了研究。到现在为止，论文已经有一万多篇，有的人说有两万篇了，但是问题没解决。到底暗能量和暗物质是什么，大家想了很多稀奇古怪的物质形态，但都不足以说服人。

物质	亮星		0.5%
	暗物质	重子	4%
		热暗物质(中微子)	0.3%
		冷暗物质(未知)	29%
	暗能量	起源于宇宙学常数 Λ	
		动力学暗能量(未知)	65%
		广义相对论在宇观尺度失效	

这个表告诉我们宇宙中物质的分布大概是什么样的。比如说，我们通常看到的亮星，根据现在的估计，质量大概占宇宙中物质的0.5%，而暗物质中的重子物质（这还不是真暗物质，这不是我刚才说的那种暗物质），包括尘埃啊，气体啊，黑洞啊，都算在内，大概占4%。还有一部分是所谓热暗物质，比如中微子，假如有质量的话，它能占宇宙中总质量的0.3%。这些加起来不到5%。这些物质是我们知道的，它们都是我们熟知的普通物质，它们的结构我们是知道的。

表中的冷暗物质才是我刚才说的那种暗物质。这种暗物质大概占宇宙中物质总量的29%。它不参加电磁作用，对光是透明的，产生万有引力，只产生万有引力。还有暗能量，大约占宇宙物质总量的65%，也是对光透明的，但是呢，它起排斥作用，因为它的压强是负的，使宇宙加速膨胀。

现在对暗能量的猜测有几种，一种认为它是动力学暗能量，就是说它是一种新的物质形态，许多搞粒子物理的人想了很多稀奇古怪的模型

来描述暗能量，论文发表了不少，但问题还是解决不了。看起来很困难，搞不清楚是怎么回事儿。

另外一些人认为，其实这东西并不神秘，并不存在什么奇特的“暗能量”，宇宙加速膨胀其实是爱因斯坦方程中含 Λ 的宇宙项造成的。看来爱因斯坦方程还是应该含有宇宙项。

还有一些人认为，宇宙项这一项太简单，只是一个常数 Λ 乘上一个 $g_{\mu\nu}$ ，仅凭宇宙项这个简单的东西，不足以解释宇宙的加速膨胀。他们认为可能要对爱因斯坦方程做大的修改。在宇观尺度上，广义相对论需要做大的修改。

如果只考虑宇宙项就能完全解释宇宙的加速膨胀，那是最好的，也是最简单的解决方式。不需要设想怪异的物质形态，广义相对论也基本不用修改，爱因斯坦方程还可以用，只须加入宇宙项。要是这样子的话，那就表明爱因斯坦所认为的他犯的最大的错误其实不是错误。对于暗能量和暗物质，现在我们所知道的情况就是这样。

我们知道，自然科学从哥白尼到现在才五百年，我们对宇宙，特别是早期宇宙的了解还是十分不够的。我们现在可以肯定的是什么呢？我们的宇宙是均匀的，是无边的，是膨胀的，是不断演化的。这些认识大概大家都还比较一致，至少现在的学术界是比较一致的。至于宇宙是有限无边呢，还是无限无边呢？它膨胀以后会不会缩回来呢？这些事情现在还是没有比较牢靠的、比较一致的看法。另外呢，对于宇宙早期演化的详细描述，往往基础不牢。因为你想，现在自然科学才诞生五百年，你想把宇宙最早期都给描写清楚，实在是根本不可能的事情。所以，对越早期宇宙的描述，越不可靠。随着科学的发展，会不断地把它完善起来。

有一位天体物理学家就跟大家说：“千万不要去追一辆公共汽车、一个女人，或者一个宇宙学的新理论，因为用不了多久你就会等到下一个。”

今天的讲座到此为止。有的网友提出来想看到一些文字资料。下面这几本书有我所讲的这些内容的文字资料。

赵峥．物理学与人类文明十六讲 [M]．北京：高等教育出版社，2008

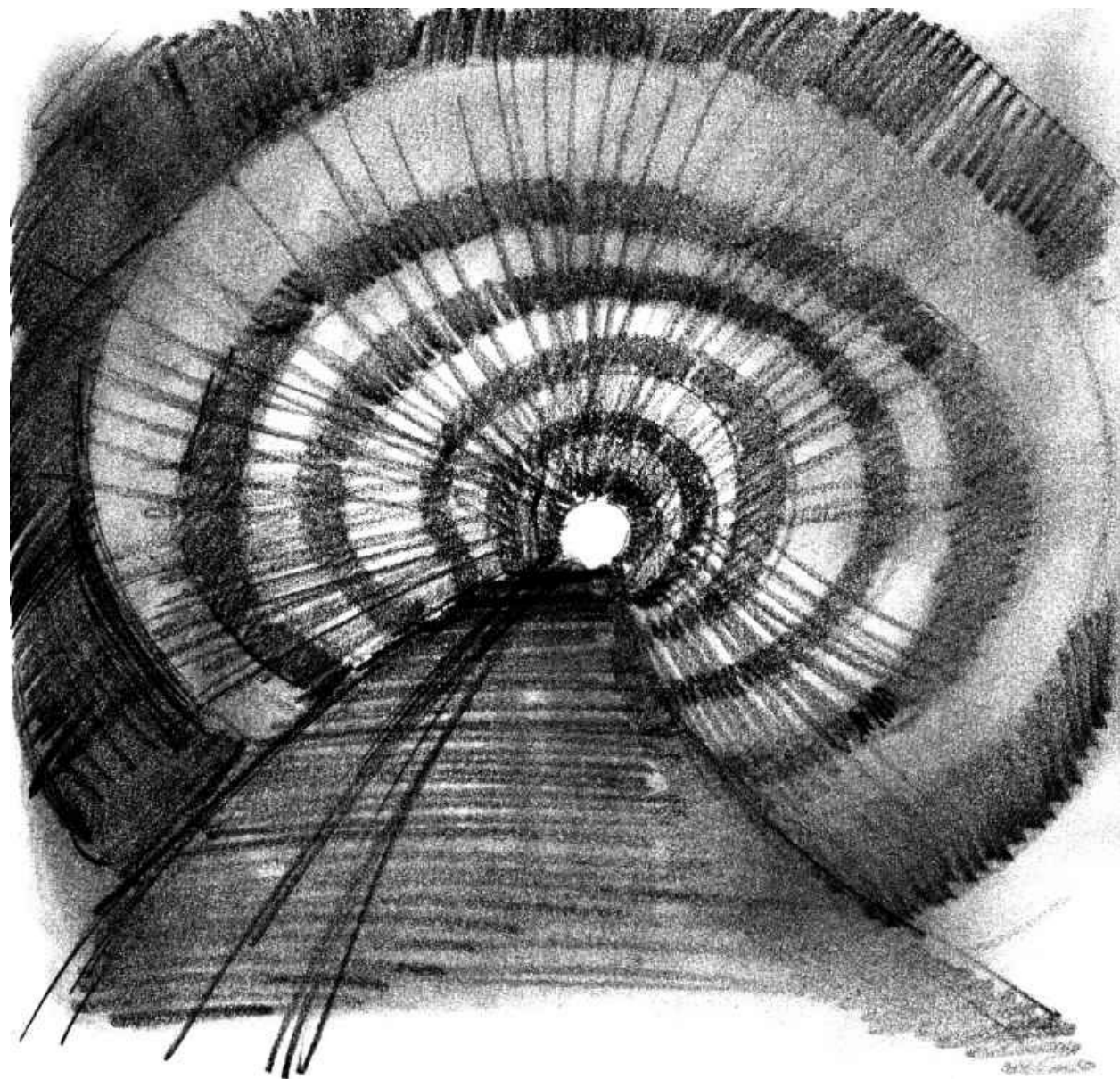
赵峥．探求上帝的秘密 [M]．北京：北京师范大学出版社，2009

新版更名为“探求宇宙的秘密”，2013

赵峥．相对论百问 [M]．北京：北京师范大学出版社，2011

谢谢大家！

第六讲 时空隧道与时间机器



绘画：张京

现在来讲第六讲《时空隧道与时间机器》，这些东西都是大家感兴趣的。我这里讲的不是小说，而是从科学的角度来探讨这些内容。不过在讲之前，我还想对“大爆炸”宇宙再做一些解释，原因是关于膨胀宇宙的内容太多，我们不可能用一讲的时间将内容都讲完，所以我现在要把上一讲遗留的一个很重要的部分，也就是关于大爆炸的一些模糊观念和认识，给大家讲一下。我讲述的这些内容发表在《科学的美国人》上面，是美国两个天体物理学家写的，紫金山天文台的陆埏先生介绍我看这篇文章，我看了以后，觉得澄清了我头脑当中的很多错误观念。陆先生也认为这篇文章很好，所以我们现在就把这部分内容介绍给大家。

1. 澄清对“大爆炸宇宙”的几个模糊认识

大爆炸没有中心

首先，“大爆炸”描述的宇宙膨胀（图6-1），是一种什么种类的膨胀？一般人可能会想，这个大爆炸模型肯定是空间当中，有一包炸药或者是什么东西“咚”一下炸了。像图6-2（a）这张图似的，有一包炸药，“咚”，炸了，刚开始，物质都集中在爆炸处，爆炸处的压强大，外面的压强小，然后爆炸物质逐渐扩散开来。显示出爆炸有一个中心。

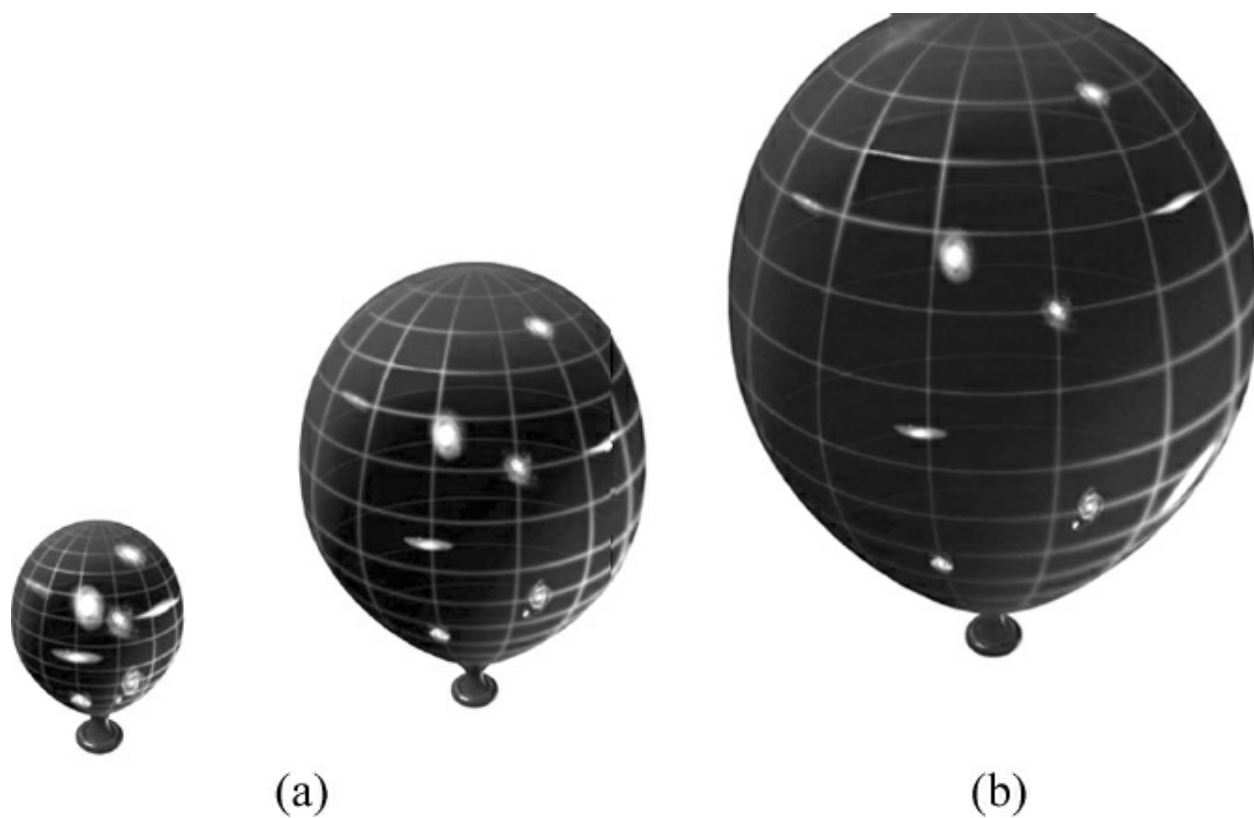


图6-1 膨胀的宇宙

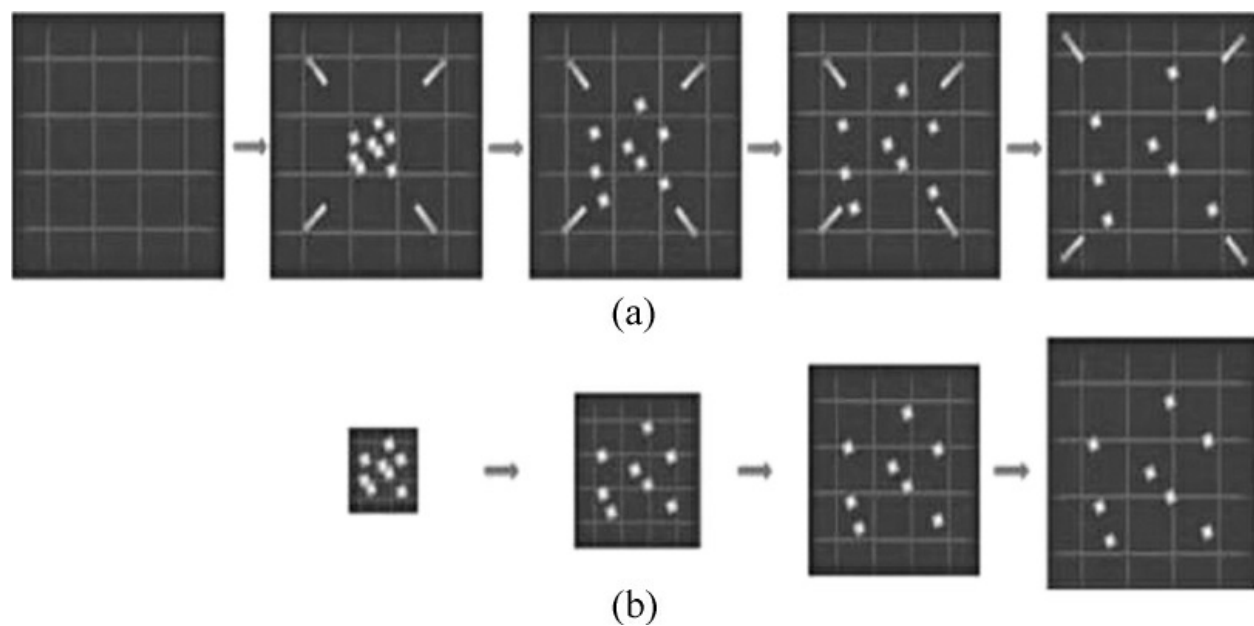


图6-2 大爆炸是什么种类的爆炸

但是，正确的膨胀宇宙模型不是这样的，按照正确的膨胀模型（图

6-2 (b))，宇宙刚开始起源于奇点，其实奇点并不属于时空。按照现在的看法，“奇点”就是时空当中抠掉的一个点，实际上宇宙是“无中生有”的。时间、空间和物质是同时诞生的，它们从奇点开始，或者说是从虚无当中产生出来的。然后你看，产生的物质是均匀分布的，空间在膨胀，物质的密度在减小，物质跟着空间膨胀散开，空间膨胀得大一点，物质就散得开一点。物质的密度随着空间的膨胀在减小，但每一确定时刻空间各处的密度总是保持相同。爆炸没有中心，或者说每一点都是爆炸的中心。

时间与物质一起从虚无中诞生

时间和空间是跟物质一起创生出来的。大家会想，科学家这想法很聪明啊。

我要跟大家讲，时间跟物质同时诞生，不是科学家首先提出来的，而是神学家首先提出来的。公元4世纪的时候，也就是公元300多年的时候，有一位著名的基督教思想家，圣奥古斯丁就提出来说，上帝在创造宇宙的同时创造了时间，上帝是在时间之外的。

他为什么想出这么一个观点呢，主要是有一帮反教会的人，在那故意提一些怪问题，让那些主教、神父们答不出来。比如有人就问了，说上帝在创造宇宙之前，他在干什么呢？这个问题不好回答。有些神职人员答不出来，十分气恼，就对提问题的人说：“上帝为敢于问这类问题的人准备好了地狱。”然而这种恐吓并不能使提问题的人服气，于是圣奥古斯丁出来说，根本就没有“以前”，宇宙诞生“以前”根本没有时间。上帝是在时间之外的，时间是和宇宙一起被上帝创造出来的，你就别问“以前”了，“以前”什么都没有。

现在的科学家也继承了这个观点，就是时空是和物质一起“同时”诞生的，在此之前没有时间。

宇宙学红移不是多普勒效应

还有一个问题，宇宙学红移是多普勒效应吗？我一直认为是多普勒

效应。我小时候看到的那些天文科普书以及后来看到的一些天文学书籍都在讲，这是一种多普勒效应，从哈勃开始就认为是多普勒效应。

大概最近十几二十年吧，我开始注意到有一些天文学的书，以及与相对论有关的天体物理的书上面，不再讲它是多普勒效应，但是也没说它不是多普勒效应。我当时想，这是怎么回事呢，说一句它是多普勒效应多直观啊，多好理解啊，他们为什么不讲呢。我估计就是写的人也没有搞太清楚，看到外国人不说了，他也不说了，但是也不知道该不该说，不能肯定。

其实呢，真正的宇宙学红移不是多普勒效应。但是我们看到的天体红移不都是宇宙学红移，实际上有两种情况，一种情况是我们星系群内部的这些银河系，它们之间相对移动，在空间中作相对运动，我们看到有蓝移有红移，这种情况确实是多普勒效应。

但是在我们本星系群之外的其他的那些星系团和星系群，统统产生红移，统统在远离我们，这就是宇宙膨胀造成的宇宙学红移，遵从哈勃定律。这种红移不是多普勒效应，而是宇宙空间自身膨胀造成的。

多普勒效应是空间本身不变化，观测者和光源在空间中作相对运动造成的。或者是光源相对观测者运动，或者是观测者相对光源运动，总之是在空间中作相对运动，这种情况是多普勒效应。你看图6-3（a）这三个图，宇宙空间不变化，星系在远离我们。右边是我们的地球，星系在向左运动，我们看到的景象是星系在退行，产生红移。但是这种效应是各向异性的，为什么呢，假如星系的左边还有一个地球，那里的观测者觉得星系在向他靠近，他看到的就应该是蓝移了。

但是我们看到四面的星系，除去本星系群内部的几个星系之外，都是红移，怎么回事呢，这是因为宇宙学红移的本质是空间在膨胀，但是星系和地球，在这个空间中的位置（坐标）并没有动（图6-3（b）的三个图）。

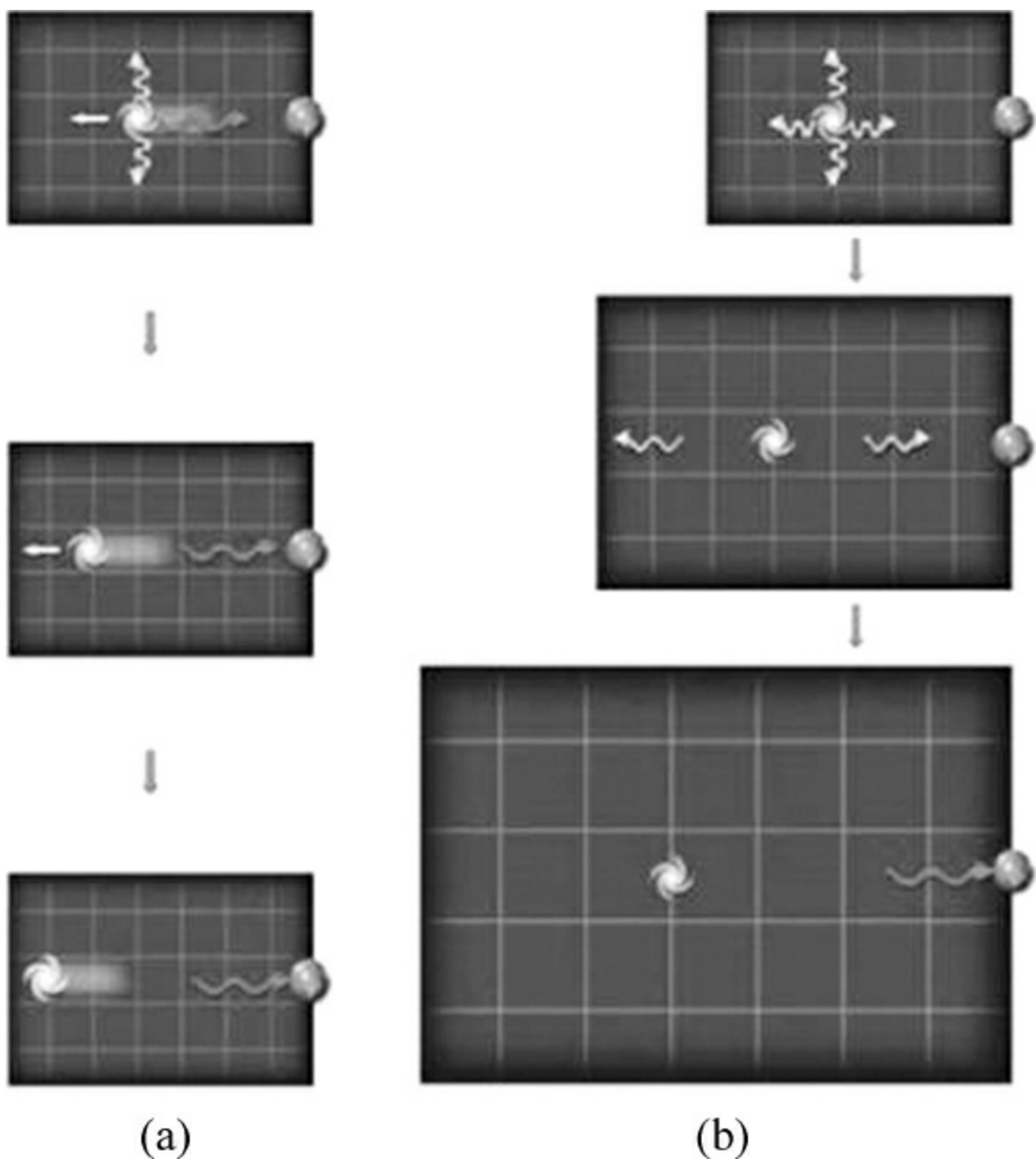


图6-3 宇宙学红移不是多普勒效应

什么意思呢，设想空间有很多网格，有坐标格，很多坐标线打的网格。星系在这组网格上的坐标点和地球在这组网格上的坐标点本身都没有变化，并不移动，并不是说跑到另外一个格点那里去了，而是这些格子本身在扩展。这样的话，星系与地球在远离，但它们不是在空间中作相对运动，而是空间本身在膨胀。这样造成的红移不是多普勒效应，

所有观测者都会认为远方的星系在远离自己，都会看到红移。这就是宇宙学红移的本质。

星系的退行速度可以超光速

宇宙学红移反映河外星系在退行，它们在远离我们，远离的速度可以超光速吗？如果它是多普勒效应的话，肯定不能超光速。为什么呢，因为相对论禁止超光速运动发生，如果一个物体在空间中作超光速运动，或者信号超光速，因果性就要发生混乱。你看一下图6-4（a）这两张图，它们表示空间没有膨胀，而星系都在远离我们，离我们越远的跑得越快，这是多普勒效应。你会发现，越远的星系逃离得越快，它逃离我们的速度会逼近光速，但是不能达到光速，否则就违背相对论了。

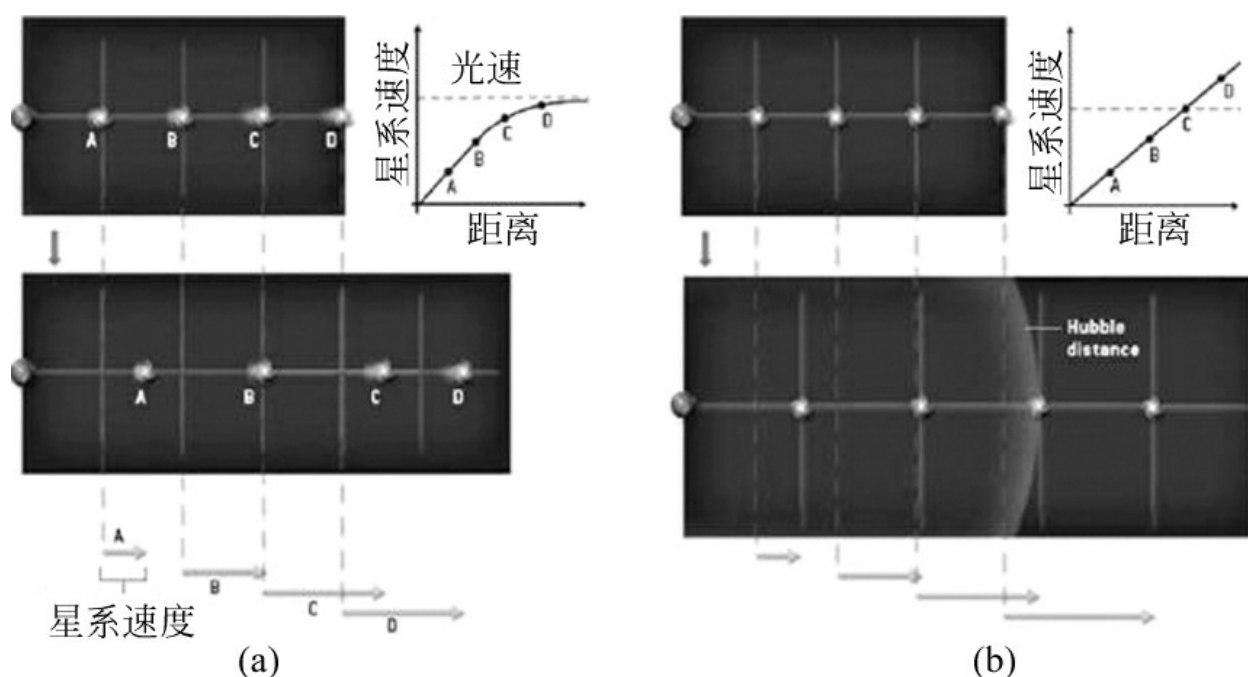


图6-4 星系的退行速度可以超光速

但是我们现在所讲的情况并不是空间不动，星系在运动，而是如图6-4（b）的两个图所示，这些星系在网格上的坐标并不动，但是空间扩展了，格子在拉大，所以看起来这些星系都在逃离我们，都在退行。离我们越远的，退行速度越快。为什么呢？你比如说，空间以10%的速度

扩张，那离我们1公里的物体，一秒钟以后距离我们1.1公里；离我们10公里的物体，一秒钟以后就离我们11公里；离我们1万公里的物体，一秒钟以后就是11000公里了，所以越远的跑得越快。

那么有一个地方，星系的逃离速度会正好是光速，这个距离被称为哈勃距离。比哈勃距离离我们更远的星系将超光速地远离我们。在哈勃距离之内的这些星系，退行速度没有达到光速。

所以远方的星系，可以逃离得超光速。我记得前些年有一位外地的老师对我说，有本书上说，河外星系的逃离速度可以超光速，我说会不会。其实我的理解是错的，他看的那本书是对的，但是当时我没有特别注意到这个问题。河外星系的退行速度是可以超光速的，它不违背相对论，因为这种退行速度不是物体的运动速度，也不能传播信号。

能看到超光速退行的星系

如果这个星系的逃离速度比光速快的话，我们还能不能看见它呢？也就是说它在哈勃距离之外，我们能不能看见它呢？回答是能看见。

为什么？本来在哈勃距离之外的星系，它那网格扩张的速度比光速快。它发射的光朝你来了，是以光速过来的，但是光子所在的网格点逃离的速度比光速还快，你应该看不见。但是这个哈勃距离又是怎么得出来的呢？大家看这个式子，

$$d = c/H \quad (6.1)$$

这是哈勃距离的公式。它来自哈勃定律 $V=HD$ ， D 是星系离我们的距离，要注意： V 是星系逃离我们的速度，如果逃离我们的速度 V 正好是光速 c ，这个距离 D 就叫做哈勃距离，哈勃距离特别用小写字母 d 来表示。哈勃距离是光速除以哈勃常数，可是天文观测表明哈勃常数不是个常数。

哈勃当年认为它是个常数，望远镜越来越好以后就发现，哈勃常数是随时间变化的，所以就改叫它哈勃参数了，只把我们今天的哈勃参数

叫作哈勃常数。哈勃参数随着时间的增大在不断地减小，这个 H 在不断地减小，所以 d 就在不断地增大，这就是说哈勃距离会增大。

你看图6-5（a）这两个图就是错误的，它表示哈勃距离是不变的。这两个图的情况，随着空间的膨胀，哈勃距离并没有变。在哈勃距离之外的这个星系，以及它射出的光子，都永远在哈勃距离之外，所以我们不可能看到这个星系，因为空间膨胀造成的它的退行速度和它射出的光子的退行速度都超过了光速，所以我们看不见它。

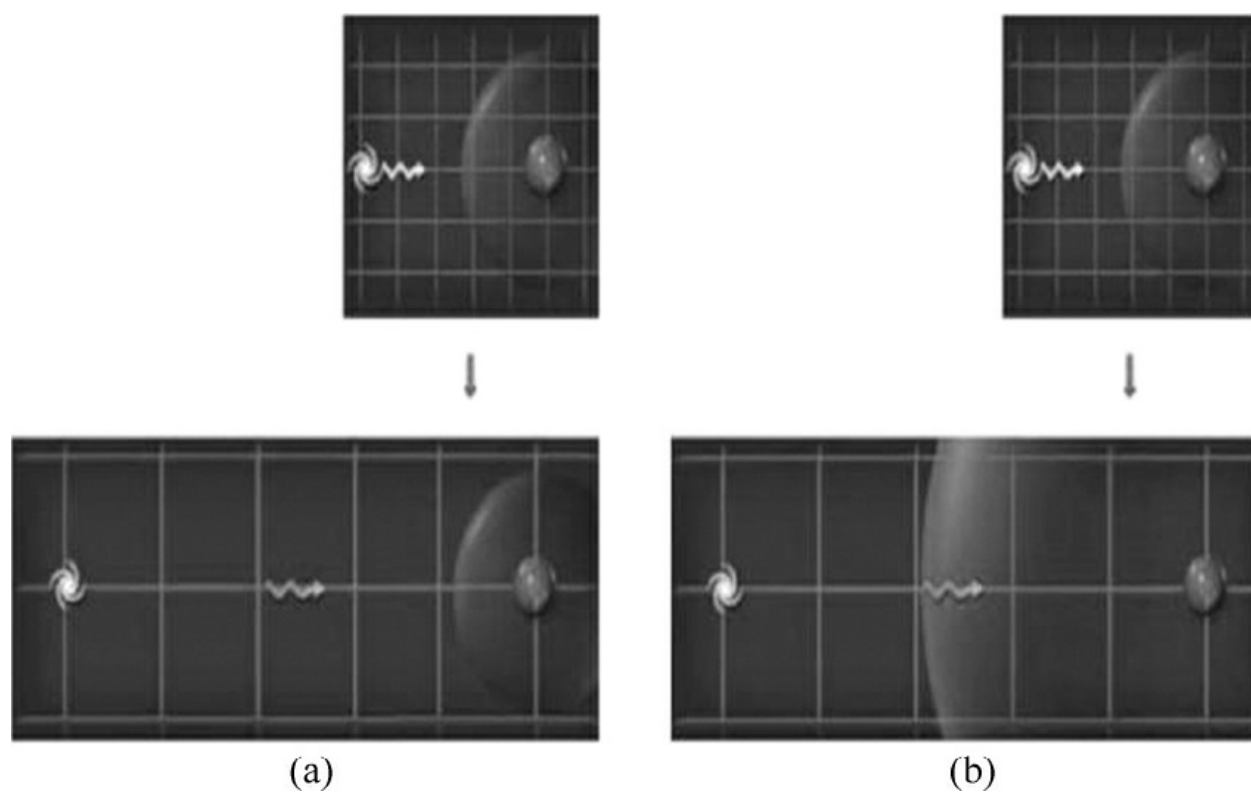


图6-5 为何能看到退行速度超光速的星系

但是请看图6-5（b）这两张图。因为哈勃参数随着时间的减小，哈勃距离在增大，最初哈勃距离比较近，后来哈勃距离变大了，光子被哈勃距离给括进来了，这个时候空间膨胀造成的光子退行速度就小于光速了，光子就能到达我们这里了，于是我们就能看见这个星系了。所以哈勃距离之外的星系发出来的光，我们还是能够看见的，这是因为哈勃距

离在不断地增大。

可观测宇宙有多大？

还有一个问题，可观测宇宙有多大？我们知道宇宙的年龄大约是140亿年，有人就想了，我们所能看到的最大距离肯定是140亿光年。我以前也是这么想的，实际上不是。为什么呢，这是因为光源发出来的光，跑过来是需要时间的。在光跑过这段距离的时候，光源又往远方跑了。设想一个发光时距离我们140亿光年的星系，我们看到它发出的光的时候，已经过了140亿年，在这段时间内，宇宙空间的膨胀，使它进一步远离我们，这个星系已经离我们不止140亿光年了。现在计算的结果认为，大概是三倍于140亿光年，也就是说，我们应该能够看到距离我们大约460亿光年的东西（图6-6（b））。现在我们看到的最远距离大概是100亿光年。这个距离说不大准，天文学上的东西，你把数量级说对了就已经很不错了，经常连数量级都说不对。由于天体离我们太遥远，你不能想象成物理实验室精密测量的那些东西，那是没法比的。这就是关于可观测宇宙的大小的问题。当然，如果空间不是膨胀的，宇宙学红移反映的是多普勒效应，那么你能只能看到140亿年前的东西，也就是距离我们140亿光年的东西。

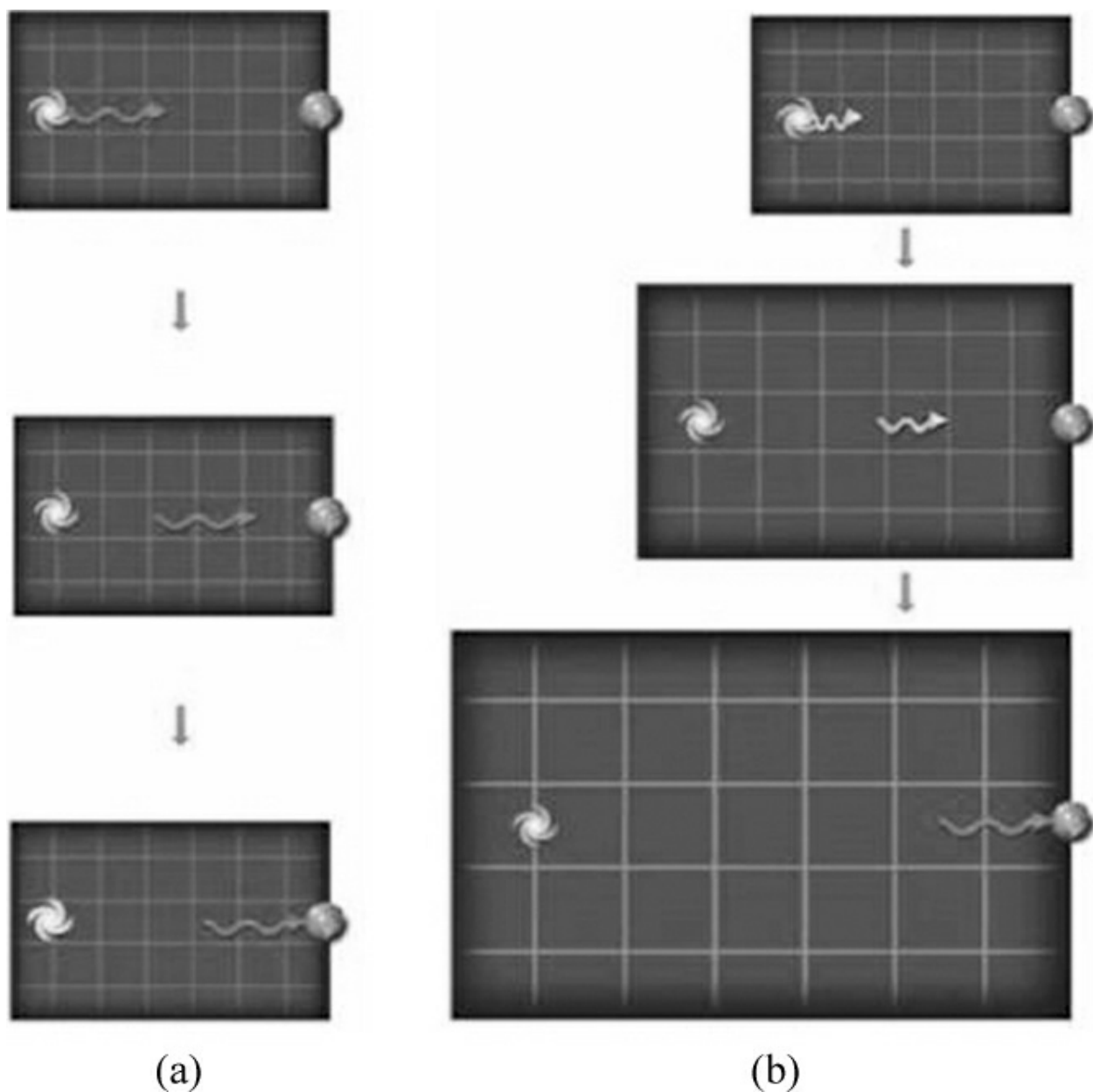


图6-6 可观测宇宙有多大

宇宙膨胀，我们自身不膨胀

还有一个问题，既然宇宙在膨胀，宇宙中所有物体是不是都在膨胀？大家想一下是不是都在膨胀呢？假如我们人也在膨胀，桌子椅子都在膨胀，尺子也在膨胀，这不等于没膨胀嘛，是不是，都在膨胀就等于没膨胀。既然我们能观测出宇宙膨胀，那我们自己和我们用的尺子肯定是后来不膨胀了，否则我们怎么可能测出宇宙在膨胀呢？

为什么我们自身后来不膨胀了呢？我们来解释一下。图6-7（a）这个图是错误的，它表示所有的东西都在膨胀，星系本身也在膨胀，要是这种情况，我们就感觉不出膨胀来了。因为你的尺子也在膨胀，那不等于没膨胀嘛。实际上是这样的，刚开始的时候，空间在膨胀，随着空间在拉大，星系本身也在膨胀，所有的东西都在膨胀。但是膨胀到一定程度以后，物质的万有引力效应（时空弯曲造成的吸引效应）就越来越发挥作用了，这个时候呢，星系团本身就不膨胀了，星系本身也不膨胀了，我们的桌椅板凳也不膨胀了，尺子也不膨胀了，但是星系团和星系团之间的距离还在拉大，还在膨胀，所以我们会看到，其他的星系团在远离我们，但是星系团本身的个头就这样大了，不再膨胀了。有人说是不是绝对不膨胀，那也不敢说，也可能还有一点，这还需要进一步深入研究。现在认为是基本上不膨胀了，比如说我们太阳系就不膨胀了，我们银河系也不膨胀了，我们银河系所在的这个本星系群也不膨胀了，其他的星系团也不膨胀了，但是星系团（群）之间的距离在拉大。

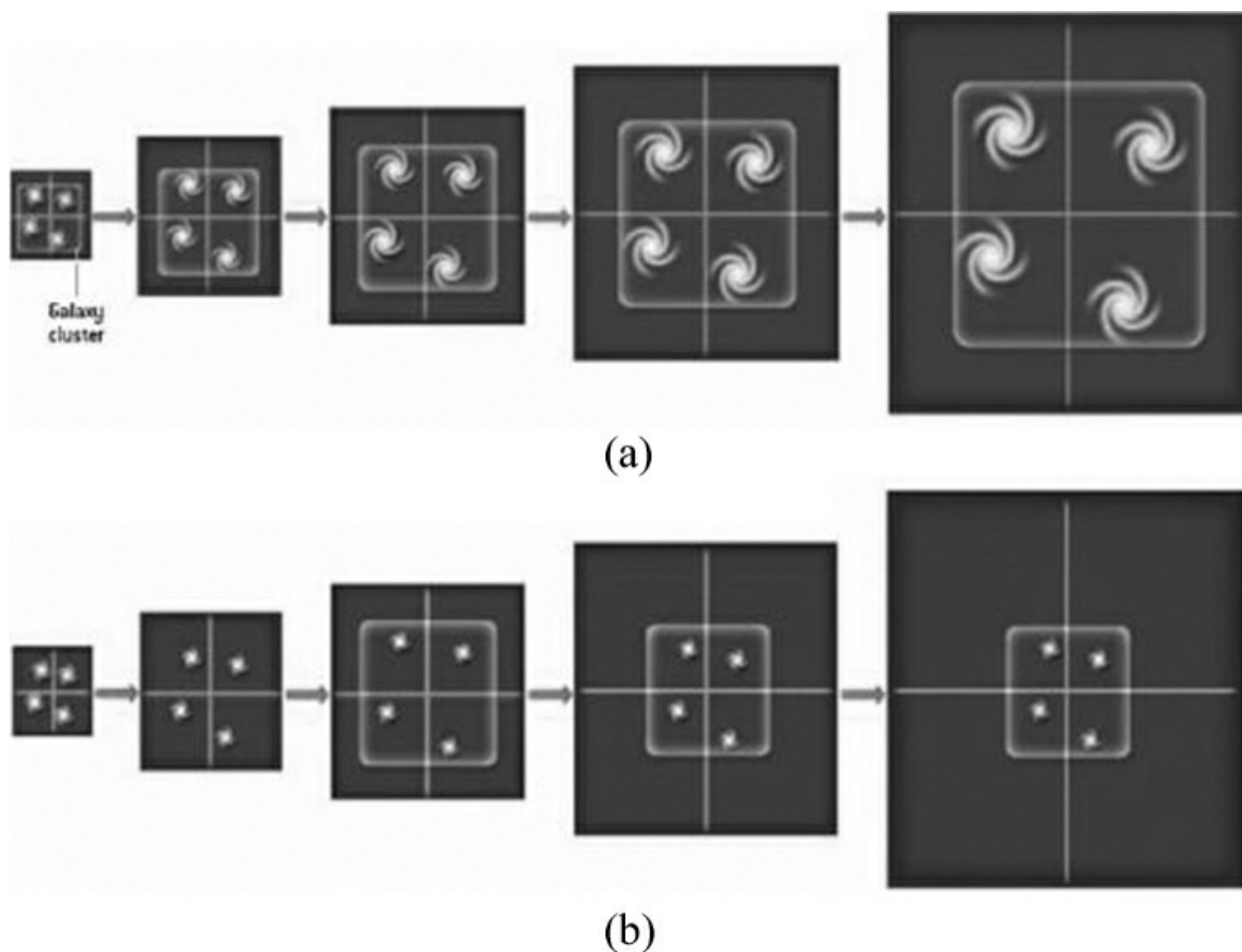


图6-7 宇宙中的星系团自身是否膨胀

2. 虫洞——时空隧道

管道与手柄

现在我们就来讲下一个问题，“虫洞”与时空隧道。什么是“虫洞”？虫洞就是时空隧道，就是连接不同宇宙的“管道”。研究表明，膨胀的宇宙可以有多个，如图6-8所示，每一个泡就代表一个宇宙，就是一个膨胀的宇宙。膨胀的宇宙可以有一些管道相通，这些管道就叫“虫洞”。有的管道的两个开口在同一个宇宙当中，就像一个手柄一样。还有一些管道，是连接不同的宇宙的。这些管道都叫做“虫洞”，或者叫做时空隧道。有了虫洞以后，时空的拓扑结构就不一样了。比如原来在我们这个宇宙泡当中，从一点运动到另一点，只可以在这个泡当中走，虽然你可

以走不同的路径，但是它们都在这个泡当中。现在呢，多了管道了，你可以通过管子走了，所以时空的结构有了一个变化，时空从单连通变成了多连通。就是可以走通过管子的路径，也可以走不通过管子的路径。

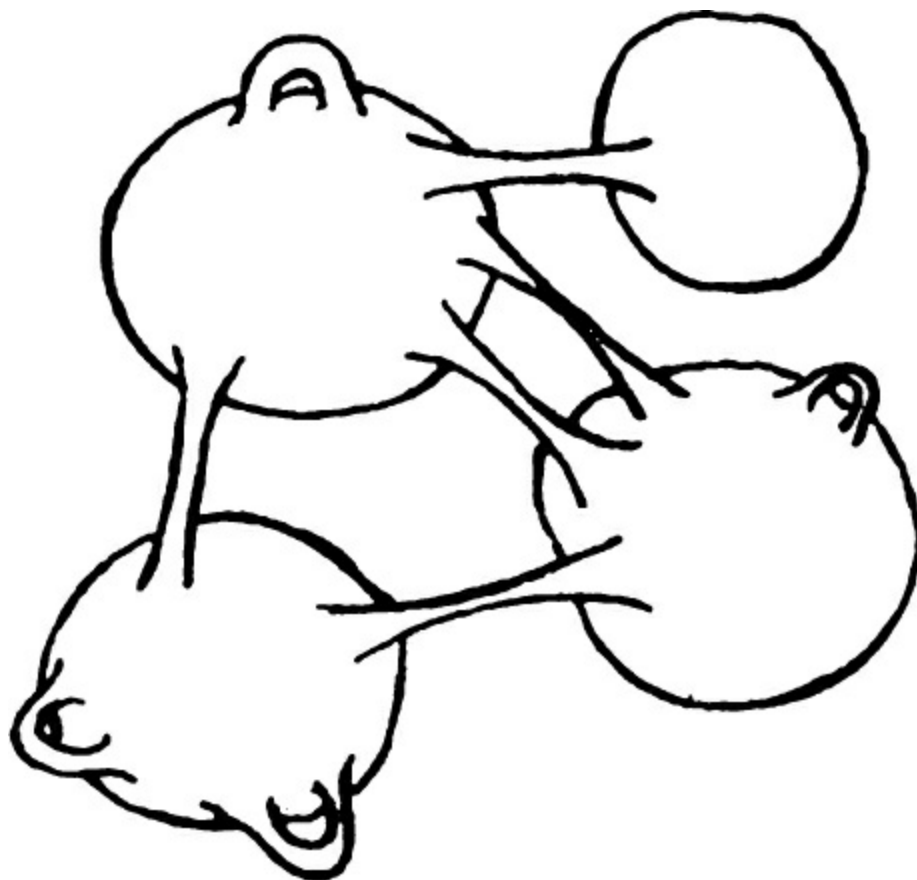


图6-8 虫洞与时空隧道：管道与手柄

现在来具体介绍虫洞，我们感兴趣的首先是可以通过的虫洞，不可通过的虫洞我也会提到。现在来说两种可以通过的虫洞。

洛伦兹虫洞

一种叫作洛伦兹虫洞，还有一种叫欧几里得虫洞。洛伦兹虫洞是能够存在一段时间的，它真是一个管子。一个洛伦兹虫洞出现的话，我们在天空中会看见一个球，这个球就是洛伦兹虫洞的洞口。火箭从这里进去，就会看到一条隧道（即虫洞），通往别的地方。这条隧道处在更高维的空间中，火箭可以通过隧道前往它的另一个洞口。虫洞的那个洞口

在哪里呢，它可能在别的宇宙当中，那你一去就很难回来了。也可能就在我们的宇宙当中，比如说它在另一个地方。例如两个洞口，一个在北京，一个在纽约，从北京进去以后从纽约出来了。此外，这种虫洞也有可能通到未来，或者通到过去。

欧几里得虫洞

还有一种虫洞，欧几里得虫洞，看不见有洞口，它是瞬时通过的。就是说物体穿过它时用的是虚时间，不是通常的实时间。比如这种虫洞口从这里飘过去碰到一个同学，这个同学就没了，没了到哪去了呢，它不需要任何时间，就在另外一个地方冒出来了，比如纽约的那个大楼上面，“噎”一下出来一人。穿过它的人还可能回到过去，一下回到大禹治水的时候，大禹身边一下“噎”冒出一个人，大禹一看，哟，怎么来了一个戴眼镜的，哪来的呢。这种离奇的情况似乎都可能产生。不过，现在的研究比较悲观，认为这种欧几里得虫洞即使存在，也不会有太大的。大概通过基本粒子还可以，太大的，通过人，现在认为可能性不大了。但是那种能长时间存在的洛伦兹虫洞，大家还都在讨论，也许有大的，人和火箭可以穿过的。

爱因斯坦的贡献——爱因斯坦—罗森桥

说起虫洞的由来，最早还是爱因斯坦提出的。爱因斯坦1915年提出广义相对论，1916年的下半年，史瓦兹希尔德，就得到了广义相对论的第一个严格解，这个解表示一个不随时间变化的球对称的时空是怎么弯曲的。这个解叫史瓦西解，最简单的黑洞就是这种时空里面的黑洞，叫“史瓦西黑洞”。

史瓦西解出来以后不久，就发现那里面不仅有黑洞，而且有虫洞。当时不叫虫洞，叫爱因斯坦—罗森桥，是爱因斯坦跟他的助手罗森搞的，它是怎样的呢。你看图6-9，这就是爱因斯坦—罗森桥。上面这个片（曲面）是一个宇宙，下面这个片是另一个宇宙。大家注意，宇宙指的是图中上、下两个片组成的曲面，指的是片，中间的空档不是，空档

是更高维的空间了。这个模型其实应该转九十度，它转九十度以后，左面是一个宇宙，右面是另一个宇宙，这两个宇宙之间，有一个喉咙通过，这个喉咙就是爱因斯坦—罗森桥，又叫“喉”。通过喉可以从一个宇宙前往另一个宇宙。但是很遗憾，研究表明，只有超光速的东西才能过得去。我们知道，超光速的东西和信号都是不存在的，所以爱因斯坦—罗森桥是不可穿越的虫洞。1935年提出来以后，有一些数学家在那里研究，搞物理的人兴趣不大。一看过不去，物理兴趣就小了。

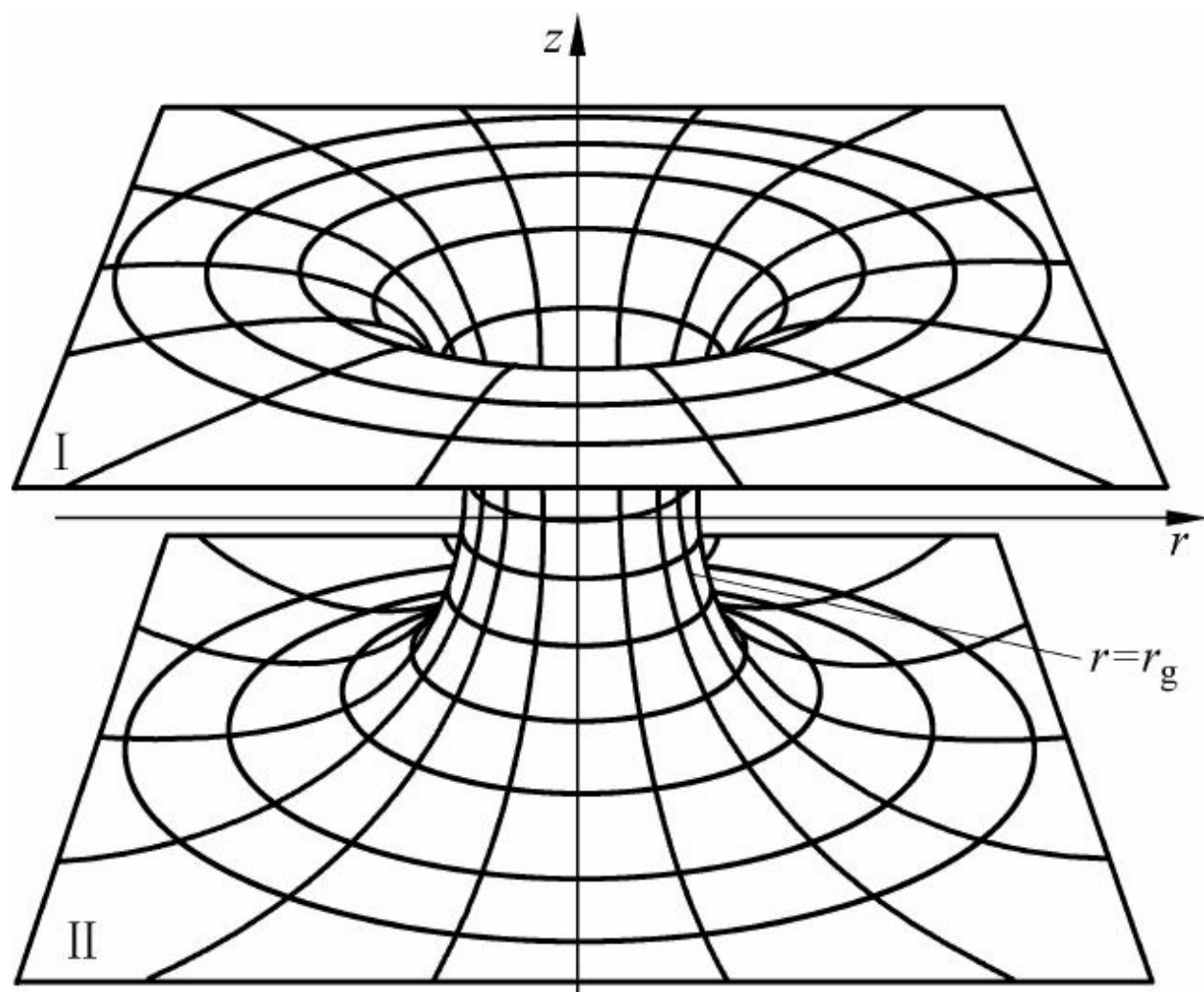


图6-9 爱因斯坦—罗森桥

没有质量的质量，没有电荷的电荷

1957年，米斯纳和惠勒提出“虫洞”这个名称，但是他们研究的虫洞

（见图6-10），仍然是不可通过的虫洞，仍然要超光速的东西才能通过。他们把虫洞的洞口看成质量，或者看成电荷，叫“没有质量的质量”，“没有电荷的电荷”。在20世纪五六十年代的时候，他们的理论还风行了一阵，我六十年代还买了一本他们写的英文书，里面就有虫洞，在那慢慢看，当时看不懂。

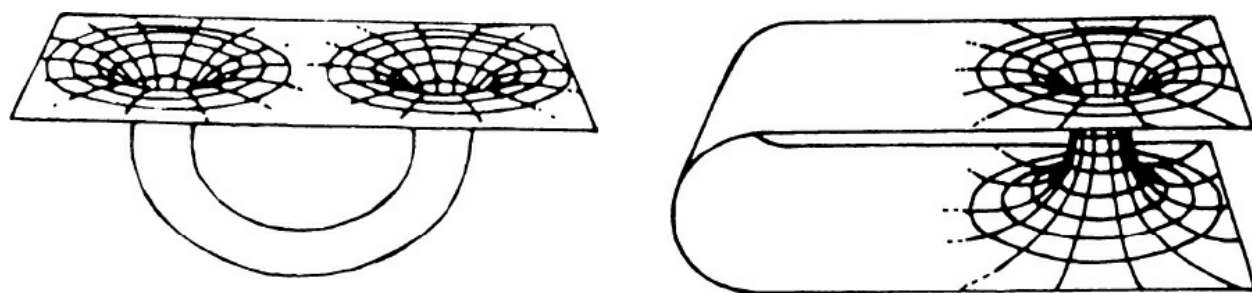


图6-10 米斯纳与惠勒提出的虫洞

这位惠勒，原来是搞氢弹的，是跟泰勒一起搞氢弹的。最早搞黑洞的一些著名人物，都是搞原子弹、氢弹的。奥本海默是原子弹的设计师，他提出黑洞的概念，当时叫“暗星”，惠勒给“暗星”起了个名字，叫“黑洞”，这个名字就叫下来了。“虫洞”也是惠勒起的名字，也叫下来了。苏联最早研究黑洞的泽尔多维奇，也是搞氢弹的，后来搞黑洞，搞宇宙学。米斯纳是惠勒的学生，他们和索恩一起，在20世纪50年代合写过一本巨著《引力（Gravitation）》。这是一本影响很大的广义相对论百科全书。

3. 可通过的虫洞

黑洞作为星际航行通道的猜想

真正提出可以穿越的虫洞，并对这类问题进行科学研究，是从1985年开始的。当时有一个天文学家叫萨根，他写了一本小说叫做《接触》，讲述人类通过时空隧道到织女星去旅行。他是这样想的，你看图6-11，时空是弯曲的，上面是我们的地球，下面是织女星，织女星距离地球26光年，就是光要走26年才能到达。

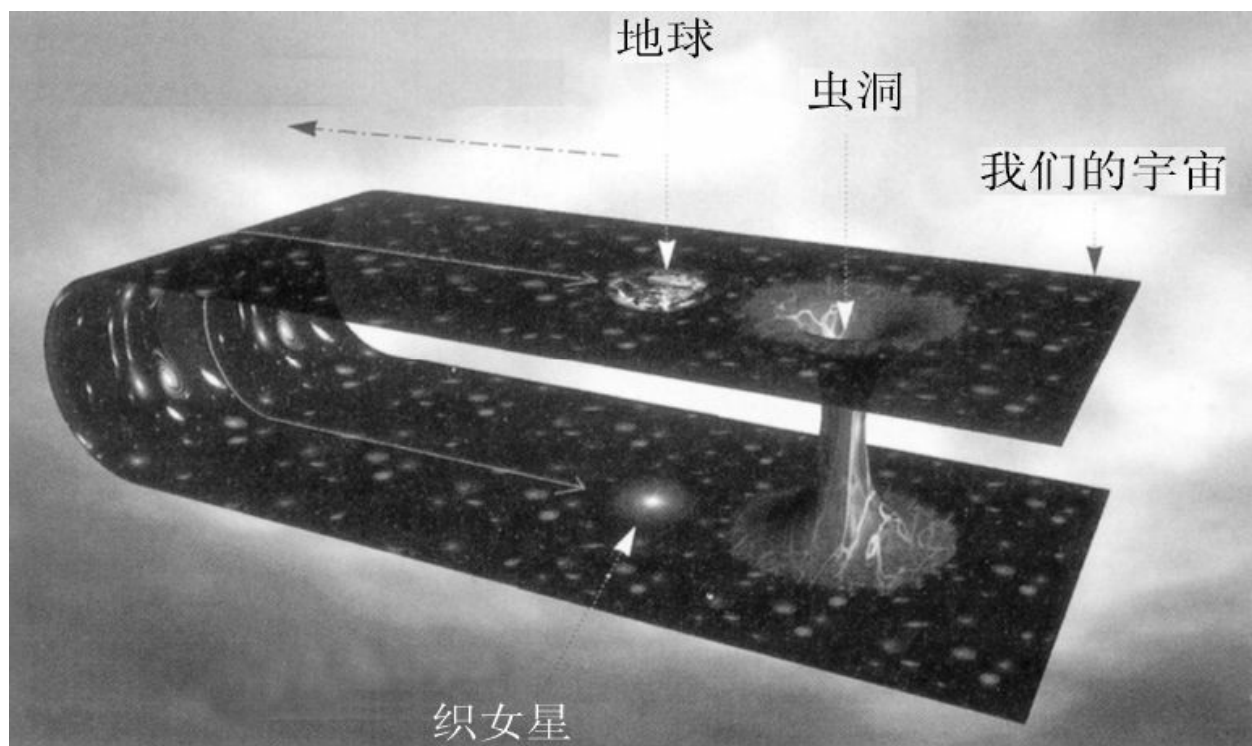


图6-11 通过虫洞做星际旅行

这个小说很难写啊，你说我发一个光信号过去26年，回来又26年，这个小说就没法写了，是不是。他说没有关系，在地球附近有一个黑洞，有个黑洞的洞口，织女星附近有个白洞的洞口。从黑洞掉进去，从白洞出来，一个钟头就穿过去了。几个小时火箭就从地球飞到织女星了，这样小说就可以写了。通过连接黑洞和白洞的管道，这个旅行就可以实现了（图6-12）。

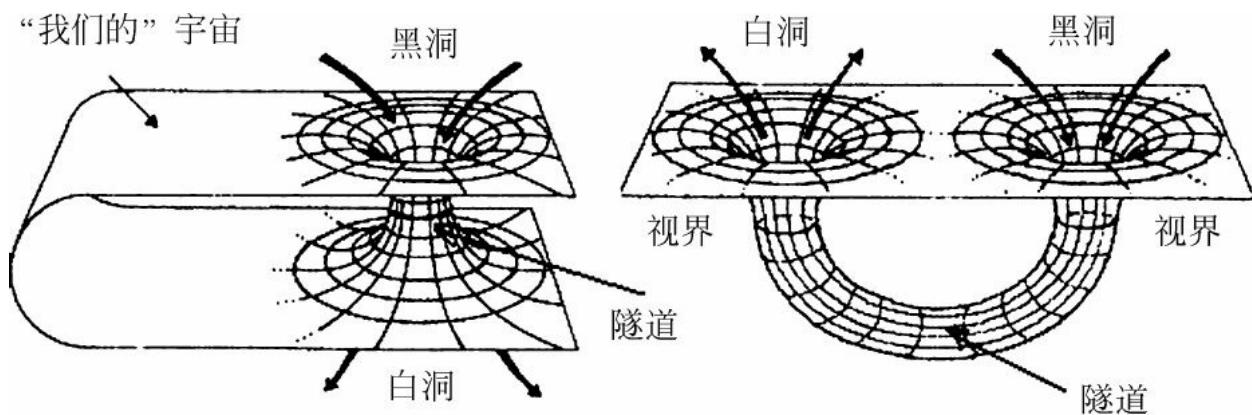


图6-12 黑洞与白洞作为时空隧道的猜想

萨根写完以后，没有把握。因为他不懂广义相对论，他是个搞行星的天文学家。大家知道，一般搞天文的人都不懂广义相对论，只有很少量的人懂广义相对论，其他人，一般也就知道一点，但是知道得不多，没有什么把握。这就跟搞广义相对论的人一样，一般的也是对天文，可能知道一点，但是知道得不多。就像我这样，就属于知道一点，但是知道不多的人。真正既懂广义相对论，又懂天文的人，还真是很少的。

萨根既然没有把握，他就写信问他的朋友索恩。索恩是惠勒的学生，上面提过的那个米斯纳也是惠勒的学生。

索恩的建议

索恩认为不行。有一些黑洞，你进去以后，里面似乎有个通道可以往前走，有人还觉得，通道的出口就是白洞。最初有些人真的觉得那地方可以通过，但是后来发现这种通道不稳定，只要有飞船在那一过，一扰动，“啪”，那喉咙就掐死了，就过不去了。索恩认为，设想通道的洞口是黑洞或白洞肯定不行，黑洞和白洞内部的通道不稳定，一扰动就断掉。

索恩建议他改用虫洞，干脆就把这个通道看作虫洞，两个洞口就是虫洞的洞口，这倒还是可行的。这样一搞呢，大量的小说、电影就出来了，描写通过时空隧道，比如洛伦兹虫洞（图6-13）、欧几里得虫洞，到未来、到过去、到远方，还有制造时间机器这一类的文学作品都出来了。

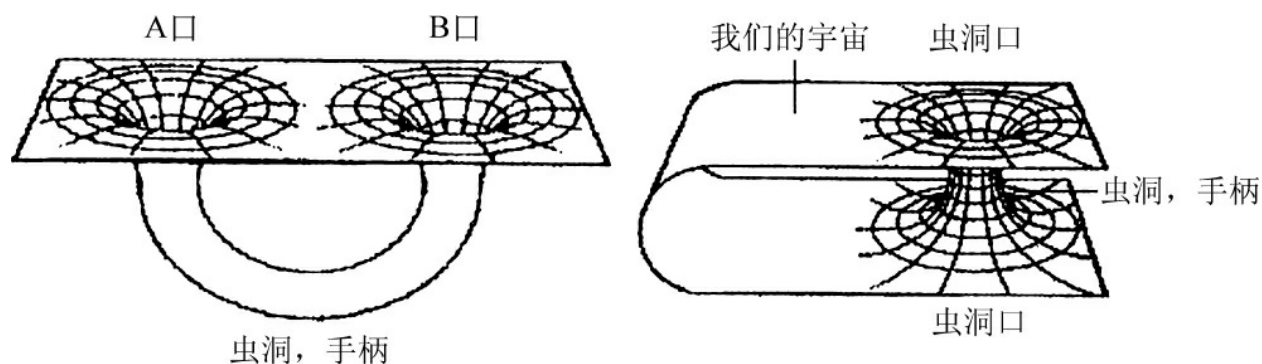


图6-13 洛伦兹虫洞作为时空隧道

我看过一个美国电影，有一个人一下子蹦到了法国大革命的时候，国王被革命群众抓到监狱里面的时候，他一下子在法国国王那里出现了，要救法国国王。这要是真给救出去了，历史不就得改写了么？这个国王路易十六最后是被送上断头机的。当时那个革命党也够恐怖的，创造了一种杀人的机器，叫断头机，把国王和王后都送上了断头机。当然电影最后的结局是没救成，要不然这事就麻烦了，历史怎么写？

索恩的开创性论文

对虫洞的真正科学研究，就是萨根的小说引起的，他使索恩开始注意虫洞问题。索恩跟毛瑞斯和Yurtsever，三个人合写了一篇文章——《时空中的虫洞及其在星际旅行中的用途》。这篇文章发表在《美国物理学杂志》上，这份杂志本来是给中学教师看的一种教学杂志，突然登了这么一篇高水平的论文，有人说使这个杂志陡然生辉。索恩的研究发表以后，霍金这些人也都进来研究了。好多人都进来研究，研究的结论是：现在的量子引力理论认为，有可能存在虫洞；改变时空拓扑、制造时间机器都是可能的，但是也不一定保证准能造出来。

量子引力的困难

所谓量子引力理论，就是把量子论和广义相对论结合的一种理论。大家都知道，量子论跟狭义相对论的结合非常成功，这种理论叫量子电动力学，后来发展为量子场论。这个理论跟实验高度地符合。但把引力

场量子化，也就是把广义相对论与量子论结合起来，这件事情碰到了意想不到的大困难。好多人进去搞，都没干出个所以然来，反正每搞一个方案，最后都发现有毛病，也不知道怎么回事。现在比较走红的是超弦理论。超弦，国内研究的人比较多，所谓比较多也就是百十个人。比如中国科技大学的卢建新，中科院理论物理所的李淼等人，还有咱们学校的吕宏。另外一种理论叫做圈量子引力。圈量子引力基本上是咱们学校一家在搞，独此一家。比如马永革教授，朱建阳教授都在做，另外还有比如说像高能所的凌意教授，他是我们这儿出去的，是梁灿彬先生的研究生，后来在国外读的博士。他们都是北师大这个组出来的。主要是这两种理论，但是遇到的困难都很大，我觉得短期内大概前景都不一定乐观，但是他们很顽强，还在那儿做。这项研究需要很深的数学，很难懂。

量子引力理论是把量子论与广义相对论结合的理论。由于它遇到大的困难，现在又提出了量子宇宙学理论，作为研究宇宙的过渡性量子理论。但是这些理论，都还确定不了虫洞一定有还是一定没有。不过，因为它可能有，所以就有人在那儿进行研究。但是初步研究表明，一旦制造出虫洞来，就会改变时空拓扑。霍金认为这就必定会出现闭合类时线。

回到过去的闭合类时线

什么叫类时线？就是四维时空中的一种曲线。比如说你们各位在位子上坐着，在三维空间当中，上下、前后、左右都确定，你们每个人都是一个点。但是在四维空间中还有时间轴呢？你们一定会随着时间前进。因为你们空间位置不动，每个人必定描出一条平行于时间轴的直线，这条线就叫做你的世界线。假如你运动，你在空间中作匀速运动，它会是四维时空中的一根斜线，你要作变速运动就会是一条曲线。这种曲线，都叫世界线，只要是描写亚光速粒子运动的，就叫类时世界线，简称类时线。

凡是静止质量不为零的东西，比如说质子、电子或者人、火箭都沿着这类曲线走。光子是以光速运动的，描出的世界线就是类光线。如果是超光速的，就叫类空线。类空线是不能传递信号也不能有物体走的，所以大家最感兴趣的是类时线和类光线。霍金说，你要造出个虫洞的话，必定会出现闭合类时线。

闭合类时线是什么意思啊？就是说这个人沿类时线转一圈儿又回到原来的位置了。这可不是说你们有个同学出去转了一圈回来又坐在这儿了，空间位置他是转回来坐这儿了，但是时间已经不是刚才那个时刻了，所以他并没有回到四维时空的同一点，只是回到了三维空间的同一点。类时线那种世界线是四维时空当中的曲线，所谓回到原位就是他要回到四维时空的同一点，回到自己的过去。也就是说，沿这条曲线的同学要回到以前的自己，这种曲线就叫闭合类时线。这种线的出现，对因果性的破坏会是很大的，一会儿我们还要再讨论这种情况。

时空的泡沫与浪花

牛顿对时间和空间做了研究。其实牛顿的很多观点来自他的老师巴罗。我们对巴罗知道得很少，巴罗是卢卡斯讲座的第一任教授。巴罗这个人很了不起，他有很多思想，牛顿关于绝对时空的很多观点是从他那儿来的。他认为牛顿比自己强多了，就把教授位置很快让给了牛顿。巴罗这个人无论从学术还是人品来讲，都很了不起。

牛顿认为时间是一条河流，一条永远不停地流逝的河流。他认为有一个绝对的空间，还有一个绝对的时间，二者互不关联。在他看来空间和时间都是平直的。爱因斯坦则把时间和空间看作是一个整体。他的狭义相对论认为时空是不可分割的，但仍认为是平直的。广义相对论进一步认为时空是弯曲的。物质的存在会造成时空弯曲。当然，没有物质的时空仍是平直的。不过，无论牛顿还是爱因斯坦，都认为时空是平滑的、光滑的。

但是从量子论的角度来研究时空的话，你就会发现时空并不是绝对

平滑的。在很小很小的范围来看，时空就不是那么平滑了，时空存在涨落，会呈现浪花与泡沫。这种情况就像海面上空飞行的飞机。当飞机在高空飞的时候，你觉得海面完全是平的，但你飞得低一点，就会看见微微的波浪，要是贴近海面去看，泡沫、浪花就全都看见了（图6-14）。

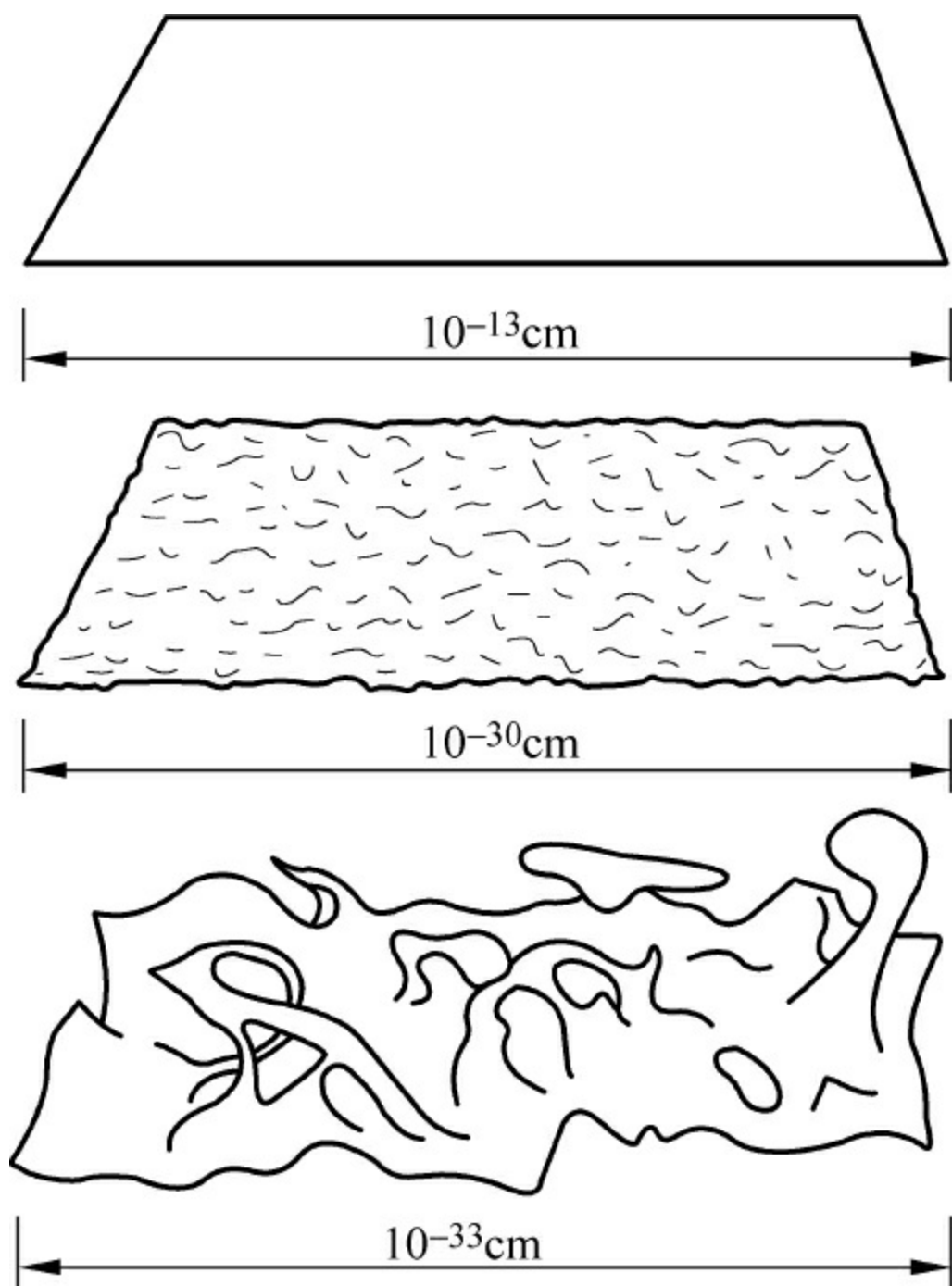


图6-14 时空如海面

宇宙泡的创生

时空也一样，并不像你想的那么平，从小范围看就会不一样。特别是在宇宙早期，整个时空处在很小范围之内，所以它肯定是不平静的。有人说，宇宙刚开始就跟一锅粥似的，处于一种混沌状态。“嘭”，冒出一个泡来，这个泡就是一个膨胀的宇宙。你看图6-15这个泡就是一个膨胀的宇宙，这中间的一块黑的，就是我们望远镜现在能看到的时空区域。“噎”，又冒出一个来。最后随着宇宙膨胀，逐渐降温以后，时空就逐渐凝固了。

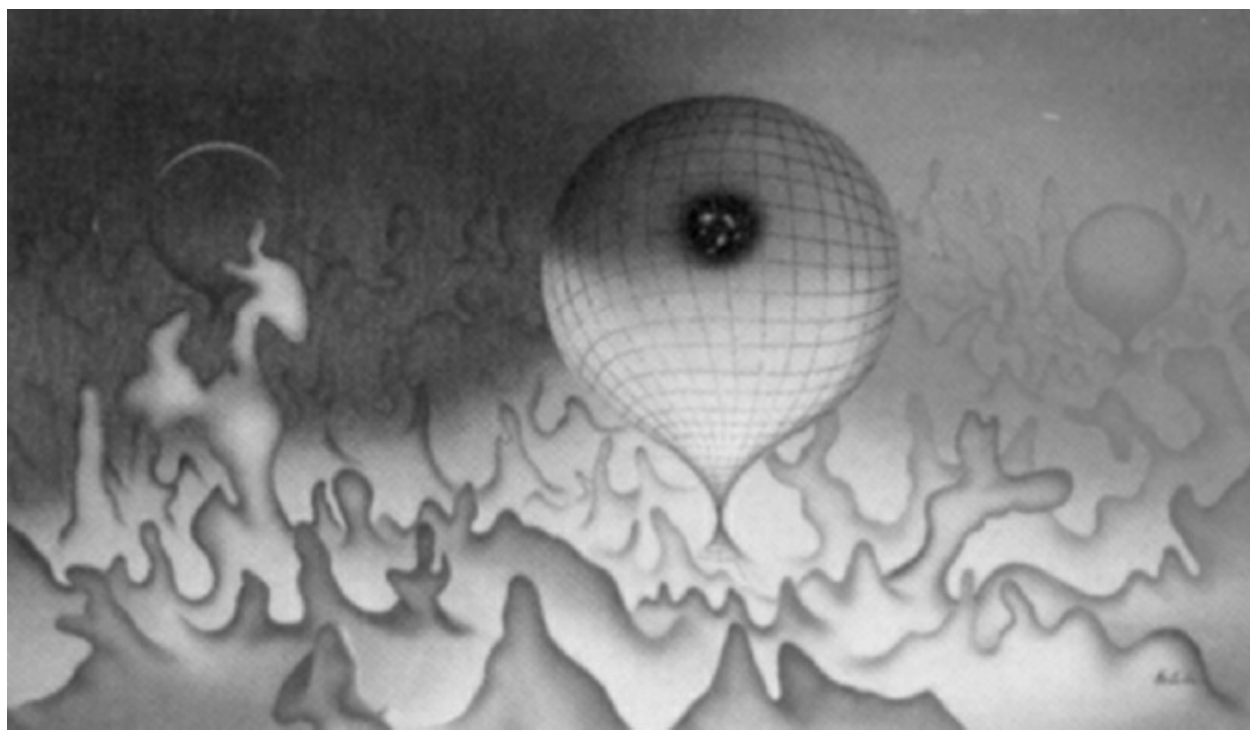


图6-15 极早期宇宙

现在对宇宙的想法也是多种多样的。比如图6-16，有人认为有一个母宇宙，有多个子宇宙，还有孙宇宙，你一看，跟癌症似的，反正想象成什么样的都有。

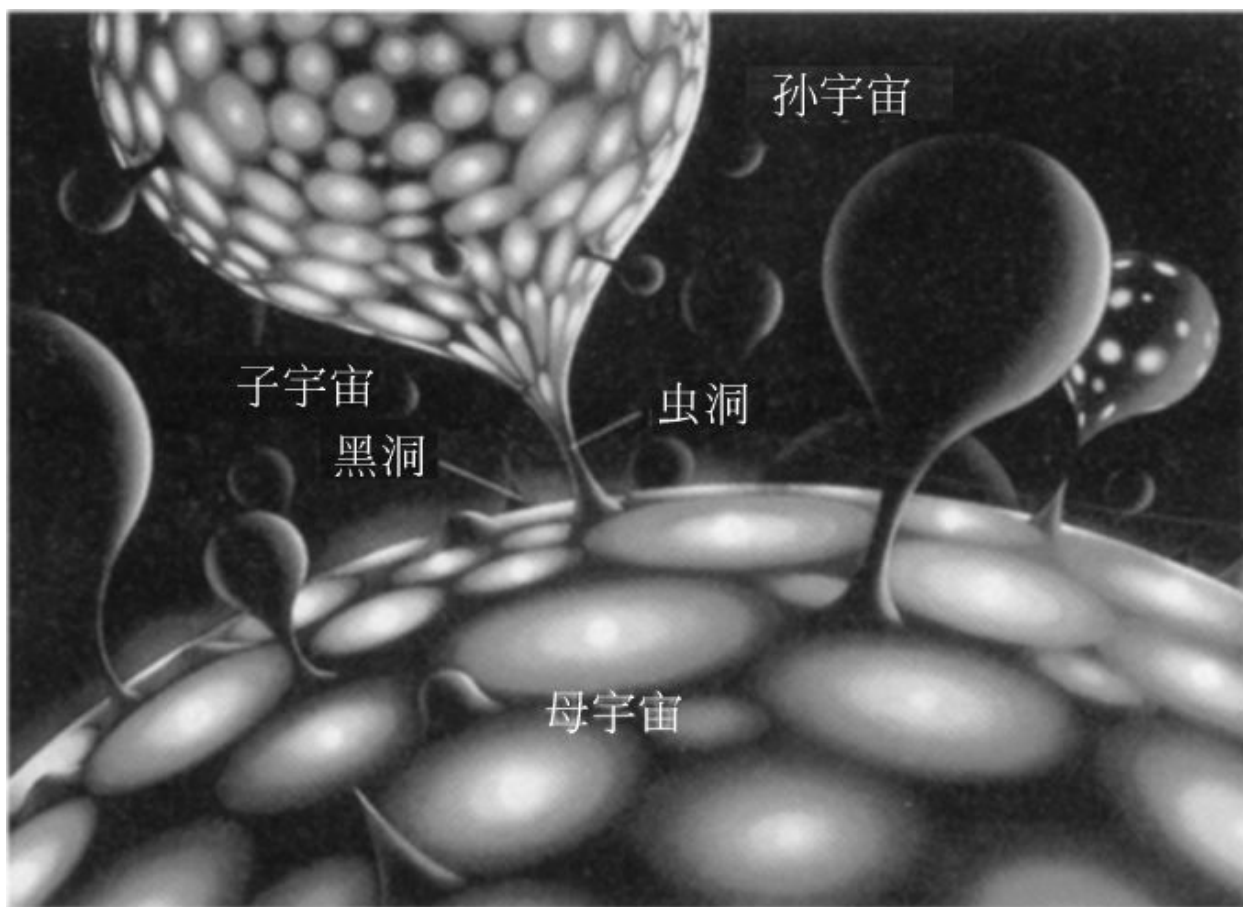


图6-16 母宇宙与子宇宙

维持虫洞需要负能量

现在，经过仔细研究发现，你要维持一个虫洞，把洞口撑开，中间还要有个喉，把喉也撑开，要想维持它的话，就需要有负能量。我们知道 $E=mc^2$ ，对于负能量，质量就是负的。对于负能的物质，按照牛顿第二定律 $F=ma$ ，你朝一个方向施力 F ，它结果不是朝运动的方向加速，而是反向加速，这就是负能物质的一个特点。负能物质咱们谁也没见过。

有人说负能物质是不是就是反物质？不是反物质，反物质是什么？就是构成原子的原子核，是由反质子和反中子组成的。反质子跟质子一样，只是带的是负电。反中子与中子的差别只在于磁矩不同。围绕这种“反核”转的是正电子，正电子与普通电子的区别仅在于它带正电荷，

那就是反物质。比如说反氢，我们已经在实验室造出来过几颗反氢的原子。反物质与正物质一样，质量都是正的。但是反物质在宇宙空间很少，当然现在有一个问题，为什么反物质会很少，这个问题还不大清楚。

负能物质也不是暗物质，暗物质的质量是正的，它是产生万有引力的。也不是暗能量，暗能量为什么有排斥效应呢，不是因为它的能量是负的，而是因为它的压强是负的，它的能量还是正的。

真空的边界效应

我们是不是绝对没有见过负能量呢，在实验室中见过——卡西米尔效应（图6-17）。卡西米尔在一九四几年提出一个观点，他说在真空当中，平行放置两块金属板，就会感觉到这两块金属板之间有一种吸引力，向中间靠。这可不是说这两块金属板带电啊，一块带正电一块带负电，那大家早就知道电荷异性相吸了，那是库仑定律决定的，根本不新鲜。卡西米尔用的这两块金属板是绝对不带电的，然而平行放置以后会产生吸引力，这种吸引力是怎么产生的呢？卡西米尔认为，这是因为这两块金属板放到真空中以后，就相当于把真空挖了两个洞，真空的拓扑结构就变化了。

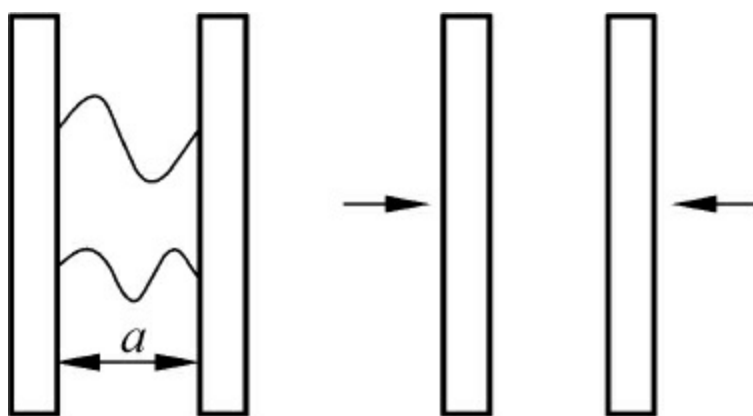


图6-17 卡西米尔效应

为什么会有这么一种往中间的吸引力呢？它是这样的，因为真空并

不是绝对的“空”、绝对的平静，真空中会不断地产生虚的粒子对，虚的正粒子和反粒子，产生又湮灭，产生又湮灭。同样的，它要产生虚光子对，在真空当中不断地有虚光子产生和湮灭。但是如果虚光子产生在这两块板之间，因为光是电磁波，电磁场在金属板上的电场强度必须为零，因此在两块板之间的虚光子就必须形成驻波。这就对虚光子的波长产生了一个限制，两板之间不是什么波长的虚光子都可以存在，只有那些波长能形成驻波的虚光子才可以存在，这就对板间虚光子的数量有了限制。而两块板外侧的真空中的虚光子，随便什么波长的都可以有，所以外面的虚光子远远多于里面的，就产生一个往里面的压力，实验上就会观测到两块板似乎有一种吸引力。

荷兰的莱顿实验室早就测到了这种吸引力，而且测量值与卡西米尔的计算值相符。现在有很多文章研究卡西米尔效应。这两块金属板之间的物质能量是负的，为什么呢？我们知道真空是能量的零点，板之外的真空就跟普通的真空一样，能量为零。这个能量零点是包括了真空涨落产生的虚粒子的贡献的。两块板中间的这个区域，因为虚光子的数量少了，所以那里的能量是低于一般真空的能量的，所以就呈现负能量。

大家看，这两块板之间的负能密度有多大。当两块板相距一米的时候，他们实验室测到的是相对于每立方米 10^{-44} 公斤这样的负能密度。这一负能密度导致两板之间产生卡西米尔力。这个密度相当于10亿亿立方米有一个基本粒子，这点负能量简直是太小了。还有没有其他负能情况呢，有，黑洞附近也有负能量。

黑洞与负能量

现在认为黑洞附近也有负能量，但是黑洞到现在一个也没确认，而且黑洞附近的负能量也很弱。可是撑开虫洞所需要的负能量实在太大了，撑开一个半径一厘米的虫洞，需要相当于地球质量的负能物质，撑开一个半径一公里的虫洞，需要相当于太阳质量的负能物质，如果要撑开一个半径一光年的虫洞，需要大于银河系发光物质总量一百倍的负能

物质。有人说有那个必要吗，有必要撑开半径一光年的虫洞吗，半径一公里不就行了吗？火箭从中间不就能飞过去了吗？不行，虫洞里面有张力，那个张力大到能把火箭扯碎，不但把火箭和人扯碎，连原子都扯碎。

虫洞作为时空通道的条件

虫洞必须达到什么程度才可通过呢？研究表明，这种张力是跟虫洞半径的平方成反比的。有人得到这么一个公式，

$$F = \frac{F_{\max}}{r^2} \quad (6.2)$$

这个 F_{\max} 是什么呢，就是物质所能承受的最大的张力，半径是以光年来计量的。如果这个虫洞的半径小于一光年的话，那个 F ，也就是虫洞里面出现的张力，会大于物质所能承受的最大的力。所谓物质能够承受的最大的力，就是原子不被扯碎的力，现在我们就以这个为标准，来研究这个虫洞是否可以穿过去。有人说那原子没扯碎，人扯碎了怎么办？人扯碎了现在先不管，现在只考虑原子不扯碎，先考虑这个问题，人不扯碎那要求就更高了。要维持这样一个虫洞，需要相当于银河系发光物质一百倍质量的负能物质。我们从来没有看到过大量的负能物质，所以撑开虫洞现在看来条件是很苛刻的。

信息穿过虫洞的奇想

但是有人产生了别的聪明想法，就是如果张三想过去，扯得太厉害了过不去怎么办呢？是不是可以把张三的信息发给那边儿的人。先给张三做一个全息的分析，然后把他的所有信息发过去。那边儿呢，因为物质的结构基本上都是普通的这些物质，然后那边再组装一个张三，这不张三就过去了吗？可是这件事情不是像你想象的那样简单，只要肉体弄个张三就行的。还有他的思想呢？他的意识呢？他的智慧呢？他的知识呢？这些东西你怎么弄过去啊，都是问题。所以这件事情，虽然说

起来好像是个办法，实际上这个办法也是非常不可行的，至少在我们可预见的将来是绝对不可能的。所以要撑开一个可通过的虫洞还真是很困难的事。

4. 时间机器

从梦想到科学

现在我们来谈一下时间机器。在相对论诞生之前就已经有人在那儿设想时间机器了。威尔斯，这个人是个很了不起的既有科学知识又有历史、人文知识的作家，他写过一本叫《时间机器》的小说，还写过不少别的书，都写得很好。1895年，狭义相对论诞生之前十年，广义相对论诞生之前二十年，他写了这本《时间机器》。这本书我没有看过，他不是依据相对论写的，那时候相对论还没有诞生。

真正出现对时间机器的研究，是在相对论诞生之后，探讨利用虫洞来造时间机器。大家看图6-18这个虫洞，它有两个洞口，一个洞口在地球上，另外一个装在火箭上，然后人坐着这个火箭出去旅行，在宇宙空间中高速地运动，然后返回来。利用狭义相对论和广义相对论结合，就能够近似算出来。结果是，这个人在十二点的时候进入虫洞口，但是呢，他还没进去，就看见自己乘坐的火箭回来了，十点钟的时候，就发现那个火箭回来了，自己从那个火箭上的另一个虫洞口出来了，他还没走就看见他自个儿从返回的火箭上下来了，回来了。研究表明，似乎能造出这么一个时间机器。

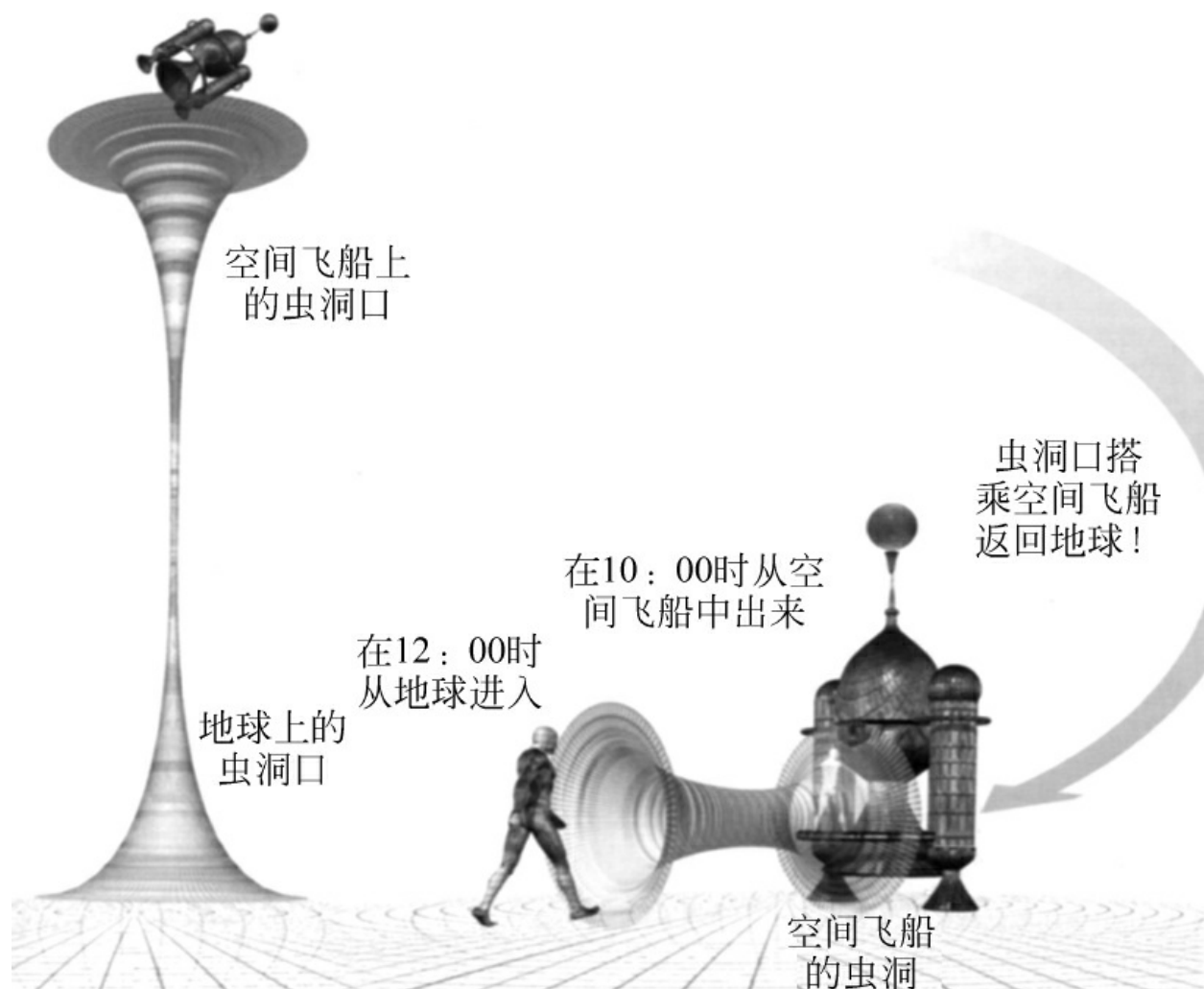


图6-18 利用飞船和虫洞制造的时间机器

在中国作者所写的书中，除去科普书以外，学术方面讲到时空隧道和时间机器的，可能只有一本书。就是刘辽先生和我，还有我们的两位博士生，张靖仪和田桂花合写的一本书，叫做《黑洞与时间的性质》（北京大学出版社，2008）。在那本书里，刘先生写了一章，是介绍虫洞和时间机器的，那本书是从学术的角度来论述的，虽然书上面有我的名字，但是我写的是黑洞部分，刘先生写的虫洞与时间机器部分。那本书是北京大学理论物理丛书当中的一本。

回到过去的时间机器

制造时间机器需要什么呢？就是要有一个洛伦兹虫洞（图6-19），

构成时间机器就一定会出现闭合类时线。要不然你也不会把它叫做时间机器。它能使一个人回到自己的过去，大家感兴趣的不就是这个问题吗？但是回到过去就构成了闭合类时线，构成闭合类时线就有很大问题，比如说这人一下回到过去了，回到他父母谈恋爱的时候，他把自己的父母给拆散了，那他是怎么出来的呢，对不对？要是这家伙是个坏小子把他妈杀了，那他更出不来了，这种问题怎么解决呢？可以说，这种问题现在还解决不了哇。

线的时空区，也就是存在时间机器管道的那个时空区。这两个区之间，还有一个叫做柯西视界的类光超曲面，上面存在闭合类光线。一般来说，需要具备这几个区域来构造时间机器。

猜想与争论

因为时间机器会破坏因果性，就像我刚才说的，所以大家就很怀疑，时间机器到底能不能造出来。霍金就认为，这个东西其实造不出来。他提出个“时序保护”猜想，说：一定有一个物理规律不允许出现闭合的类时线，不允许一个人回到自己的过去，也不准把信息传到过去。比如说这个人考大学没考上，因为题没做出来。那怎么办呢，赶紧把信息发回去，发回到他考前时候，看看这题是怎么解的，看会以后他就又考上了。那他到底考上没考上呢？实在太荒谬了，所以一定要有一个物理规律能够阻止他，不能回到过去，也不能把信息传回到过去。

另外有一位俄罗斯的物理学家诺维科夫，现在他去了美国了。这个诺维科夫也是一个不简单的人物。他提出了一个“自洽性原理”，说可以让人回到过去，但是不能破坏因果性。这叫“自洽性原理”。

诺维科夫挺有意思，刚开始他不认识萨哈罗夫。萨哈罗夫是苏联物理界的泰斗级人物，氢弹的设计师，一次，萨哈罗夫在核武器试验场工作时，到诺维科夫所在的研究所办事。萨哈罗夫这个人平常衣着不整，不大注意外表。他一推门进了诺维科夫呆的屋子，诺维科夫一看，以为萨哈罗夫是一个老百姓，找他们要啰嗦什么事儿呢，就喊：“出去！看你那个样子，把你的衣服好好整一整。”萨哈罗夫还真的就出去了，到洗手间把衣服整了一下。他们同屋的一个人，一下跳起来了，说：“你疯了，他是萨哈罗夫。”哦，原来是萨哈罗夫。不过，一会儿萨哈罗夫回来了，他也没有为这事生气，就和他们讨论起问题来了。

那么什么物理规律能够破坏闭合类时线呢，霍金认为，可能是真空极化的能量，靠近柯西面那儿有真空极化的能量，而且那个能量很大，可以把闭合类时线破坏掉，他做了一个证明。后来刘辽先生跟他的两个

研究生（李立新和许建梅）合写了一篇文章，对霍金所提出来的那个猜想，他们找了一个反例。李立新现在回国了，这个人本科是北大毕业的，后来考了我们的硕士生，在我们这儿学的相对论。他又去了美国，在普林斯顿大学拿了博士学位，然后在哈佛读了博士后，在德国的爱因斯坦研究所工作了五年，现在回到了北大，在科维里理论物理研究所工作。他现在是世界上比较有名望的搞宇宙学和时空隧道理论的专家。

那么这个阻止回到过去、影响过去的规律到底是什么呢，我自己有个猜想，也许就是热力学第二定律。热力学第二定律告诉我们，时间只能往前发展，不能够转回来。所有的自然过程都是不可逆的。我认为霍金的时序保护机制可能就是热力学第二定律。根本的物理定律不会太多，所以很可能归根结底是热力学第二定律。

物理学的两个特别分支

我们知道物理学当中有两个分支是特别值得注意的，一个是广义相对论，因为其他的物理学分支都认为时空是平直的，物质在里头相互作用，物质和时空之间相互没有影响。就好像时空是舞台，物质是演员，演员跟舞台相互没有影响。只有广义相对论认为物质的存在会使时空弯曲，也就是说演员对舞台会有影响，舞台对演员也会有影响，这是广义相对论的一个特点，除去广义相对论的所有的物理学分支都是不考虑物质对时空的影响和时空对物质的反作用的。

还有是热力学，包括广义相对论在内的所有的物理学分支都认为时间是可逆的，就是可以去又可以回来的。广义相对论里面虽然有黑洞还有白洞，但它算出来的只是个“洞”，并没有告诉你是黑洞还是白洞。黑洞和白洞都是爱因斯坦方程的解，两者是对称的，并不体现不可逆性。其他的物理分支，包括量子力学全都是可逆的。唯独热力学第二定律告诉我们，自然过程不是可逆的，时间是有一个流逝的方向的。虽然现在有人谈论热力学的时间箭头，宇宙学的时间箭头，心理学的时间箭头，我认为所有这些箭头归根结底都是热力学的时间箭头，都是热力学第二

定律的表现。为什么会有热力学第二定律？为什么会有不可逆性？不清楚。这种不可逆性，不能够从其他的理论推出来，所有从其他物理理论推导出不可逆性的企图都失败了，这个事情也是很奇怪的。

阿瘡瘡！

好，我们看，假如宇宙中真的出现了一个虫洞，大家会看到什么呢？会看到一个球状的虫洞口。那个洞里的景象跟外面的天空是不一样的。正如李白的诗：

洞天石扉，
訇然中开，
青冥浩荡不见底。

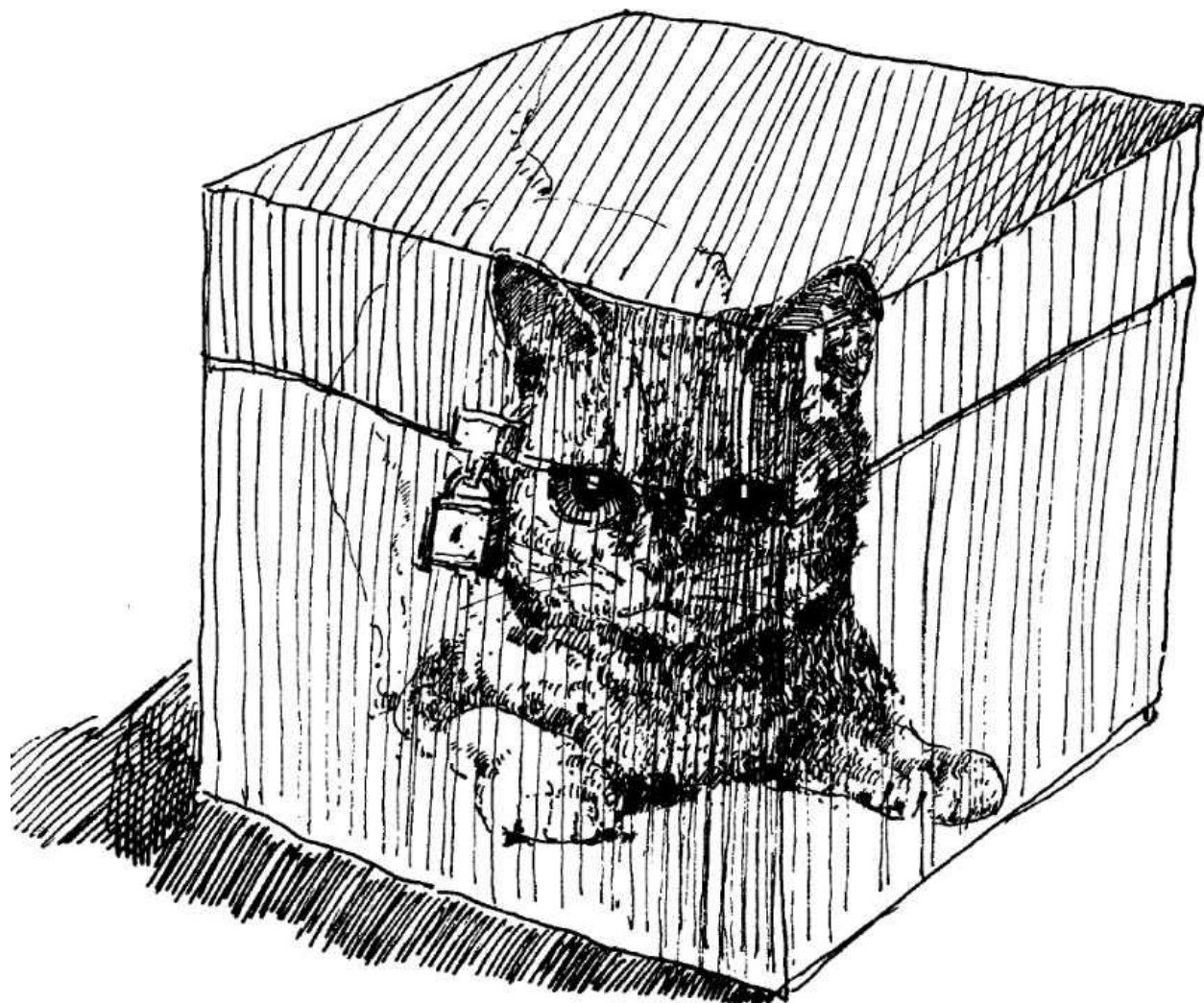
你看到的景象可以说是别有洞天了。还有一位艺术家写过这么一首诗，说：

只闻白日升天去，
不见青天降下来，
有朝一日天破了，
大家齐喊阿瘡瘡

这个球状的虫洞口，难道不像天破的一个洞口吗？“瘡瘡”两个字我问了好几个人它念什么，然后又查了些字典，最后确认它念“瘡瘡（guǎi guǎi）”，苏州那边的人表示惊讶的时候常常说：瘡瘡。这两个字，不是小乖乖的那个乖乖，而是这个瘡瘡。这首诗的作者是谁呢，是才子画家唐伯虎，是他在自己的画卷《白日升天图》上面题的一首诗。

我今天的报告就讲到这儿，谢谢大家。

第七讲 激动人心的量子物理



绘画：张京

这一讲介绍量子物理的内容。主要介绍原子物理学和量子力学的发展，以及在发展过程中的论战。大家都知道，量子力学发展中的论战是很激烈、很有趣，也很具启发性。

1. 原子物理学的发展

元素周期律的发现

我们首先要讲一下原子物理学的发展。原子论是从古希腊开始就有的，古希腊人认为，原子是物质的最小单元，是不可分的。到了19世纪，人们逐渐发现了一些新东西。首先是，1869年门捷列夫发现元素周期律。

在此之前，英国的化学家纽兰兹做了重要探索，他发现，如果把元素按照原子量大小的顺序排列起来以后，就会有一个规律，基本上是八个元素一个周期，化学性质会有一个周期性变化，他把这个规律叫做八音律。在英国皇家化学学会上，他做了一个报告，结果被大家讽刺了一通。有人说，你很聪明啊，你怎么想起来把元素按原子量的顺序排一排，你怎么不把元素按拉丁字母的名称abcd排一排，看看有什么规律啊。讽刺他。结果纽兰兹就没有再搞下去。当然，人家也抓住了他的一些弱点，他的顺序排得比较呆板，八个，八个，.....，前两行还不错，第三行以后就有点问题了。

比纽兰兹稍微晚一点，俄罗斯的化学家门捷列夫得到了一个更准确的规律。门捷列夫，有一本书上说他是俄罗斯人和蒙古人的混血儿，出生在中亚。他们家八个孩子，他是最小的。父亲去世以后，他母亲经营工厂，最后经营不下去了，于是就回到了俄罗斯的内地。他母亲非常辛苦，一个人将八个孩子带大，最后把最小的儿子送进了师范学院。这个最小的儿子就是我们熟知的化学家门捷列夫。为什么把他送进师范学院呢？因为师范学院不用交学费，他家里经济条件实在太困难了。

我想，化学家中没有比门捷列夫更杰出的人物，只有跟他水平差不多的。但是，门捷列夫既没有得诺贝尔奖，也没能当上科学院院士。为什么没当上科学院的院士呢？因为他同情学生运动，沙皇政府不允许他当院士。诺贝尔奖呢，在评奖的时候，有一次是提名了，讨论的结果是四票对五票，一票弃权，把他给否了。

那次奖给了谁呢？给了一个假发现：化学家莫瓦桑用石墨制出了金刚石。因为当时化学家已经知道石墨和金刚石都是碳元素，莫瓦桑是一位杰出的化学家，他就想用石墨烧制出金刚石。结果，烧了一炉没有，烧了一炉又没有，但他坚信一定能烧出来，就继续烧。他的助手都认为他烧不出来，但是又没法说服这个老板不干，所以在有一次装炉的时候，有一个助手就在里面搁了一颗金刚石。等到开炉的时候，发现里面有一颗金刚石，高兴坏了，烧出金刚石来了。莫瓦桑到死都不知道他没烧出金刚石来。

现在我们知道石墨是可以造出金刚石的，但是需要高压，莫瓦桑是在常压下做的。这件事情责任不在莫瓦桑，但是这个发现是个错误发现。另外呢，错误的评价使得门捷列夫没有能够得奖。

第二年，这个评委会觉得，这次奖大概应该给门捷列夫了，结果他死了。诺贝尔奖只给活人，不给死人，所以在座各位要想得诺贝尔奖，身体还得搞好一点。不但要做出成就，还要活得有挺头，挺到发奖的那一天，对吧。

伟大的师范生

门捷列夫非常杰出，我们大家看。有两位伟大的师范生：一位是最杰出的化学家门捷列夫；一位是最杰出的物理学家爱因斯坦。因为爱因斯坦上的是苏黎世工业大学的师范系，上的是一个培养大学和中学的数学、物理教师的一个专业。大家知道，还有一位非常伟大的师范生，是谁呀？毛主席。他是湖南第一师范的师范生。

所以我们师范院校的学生，应该有信心，只要你好好干，你是可以

干出最优秀的工作的。而且像我们这个学校，老师是不会耽误大家的，只要你肯干，你会是很有希望的。前些年我作报告时开玩笑，说我们北师大虽然没有培养出诺贝尔奖获得者，但是培养了一个诺贝尔奖获得者的父亲。杨振宁先生的父亲是我们北师大数学系毕业的。现在，我们北师大中文系，不是已经培养出了诺贝尔奖获得者吗？

光谱线的规律

在周期律发现前后，还有一些重要发现。一个是发现了光谱线，发现各种元素都有一根根光谱线，但是找不到规律。

有一个做光学实验的人，想起自己的一个朋友，是位中学的数学老师。这人有一个爱好，对任何自然现象都想找找有没有数学规律。于是，这位搞光学实验的人就把这个事情告诉了这位数学老师。

这位数学老师叫巴耳末。巴耳末弄来弄去就搞出来一个巴耳末系，找到了规律。第一个规律找到了，第二个规律大家就会用类似的方法找，就比较容易发现了。这些规律的发现，对玻尔轨道模型的提出，是有启发性的。所以巴耳末的贡献是很大的。大家注意，他也是一位中学老师。

X射线的发现

此外，1895年的时候，伦琴发现了X射线。别的发现，一般老百姓都不大容易知道，不大容易感兴趣。这X射线一出来，立刻轰动了当时西方有文化传播的世界。

为什么呢？别的你不懂，说这东西一照能把骨头照出来，老百姓都懂。所以大家兴趣很浓厚。有很多贵族和贵妇人都要求看看伦琴的那个装置，那个X射线，把手放在那儿照一下，看看怎么样。伦琴也不怕耽误工夫，每次都细心地准备，每次都给大家讲解。讲完以后，就说我们实验室比较缺经费。

我们现在已经知道了，光谱线是原子外层电子的能级跃迁，或者说外层轨道之间的电子跃迁产生的，是外层轨道的行为。

那么伦琴射线呢，伦琴射线也不是原子核反应，它是用电子束把原子内层电子打飞以后，外层电子跃迁过来产生出来的，能量比较大。但还不是核反应，只是核外电子的行为。

天然放射性的发现

真正的核反应是1896年贝克勒尔和居里夫妇发现的。贝克勒尔是法国的一位物理学家。他研究铀的时候，有一次把铀矿石放在了抽屉里面，抽屉里面有一堆没有感光的黑纸包着的胶片，上面有一把钥匙，他把铀矿石放在了钥匙上面。结果，在他要使用这些胶片的时候，发现胶片感光了，有这把钥匙的像。他猜测铀矿石是不是发出了什么射线呢，当时大家认为可能是X射线，或者是什么其他的穿透力强的射线。

不久居里夫妇对这一现象进行了深入研究。当时皮埃尔·居里刚刚结婚，他的夫人就是非常著名的居里夫人。那时居里夫人要做博士论文，皮埃尔建议她选了这个题目。

奋斗的女生

居里夫人是波兰人。当时波兰已亡国了，被俄罗斯和德国瓜分了，东部归了俄罗斯，西部归了德国。他们家分在了俄罗斯占领的那一半。居里夫人的父亲是一位中学老师，他们家的孩子都很用功。居里夫人的波兰名字叫玛丽娅，她年轻的时候受到过两次人格上的侮辱。第一次是沙皇派来的监督官，认为她的俄语讲得不好，把她叫起来站在那里，训了一顿。她感到非常大的羞辱，觉得自己的祖国灭亡了，才受到这样的侮辱，因此她非常爱国。

另外一件事发生在玛丽娅高中毕业以后。她本来很想继续深造，但是那时候全世界只有法国的大学招收女生。而且呢，她姐姐已经去法国了，他们家的经济条件不可能再供她上大学，她就当了家庭教师。那一年放暑假的时候，主人家在外面上大学的大儿子回来了，跟玛丽娅的年龄差不多，两人就谈上恋爱了。这一家的女主人，也就是男孩子的母亲，把儿子训斥了一顿，说你怎么能找一个平民的女儿呢，咱们家是贵

族。虽然他们都是波兰人，玛丽娅的父亲还是知识分子，是中学的教师，但他没有贵族身份，于是女主人坚决切断了这段关系。玛丽娅感到非常屈辱，非常难过，一气之下去了法国，到法国求学。

玛丽娅上大学的时候特别刻苦，大家可以看她的小女儿给她写的传记——《居里夫人传》，从这本书里可以看到她的奋斗经历。这本书写得非常好。玛丽娅当时生活很艰苦，住在一个小阁楼里，有好几次晕倒了，邻居赶紧去找她姐姐和姐夫来看她。一看，她家里什么吃的都没有，就一点儿水萝卜，严重的营养不良。后来姐姐就把她接到自己家养几天。但是她姐姐家离学校很远，她很快又跑回去住了。这样，到毕业时候，她拿了数学和物理两个硕士学位。特别幸运的是，在毕业前夕，经人介绍她认识了皮埃尔·居里。

居里夫妇的结合

皮埃尔·居里比她大好几岁。皮埃尔当时已经做出成就了，他的实验做得很好，跟他哥哥一起得出了磁学的居里点和居里定律，还发现了压电效应。皮埃尔很喜欢玛丽娅。因为当时女孩子学自然科学的很少，即使能够上大学，也基本上都是学文科的。皮埃尔希望玛丽娅毕业后能留在法国。她说不行，她想她的父母，她想回家，怎么说都不行，皮埃尔很遗憾，只得看着她走了。

走了半年多，皮埃尔突然收到了玛丽的信（玛丽是她的法文名字），说她在家的情况确实像皮埃尔对她说的：你回去什么也干不成，还是在法国能干点什么。她觉得真是这样，她说还想回法国。皮埃尔·居里一听，那简直喜出望外啊，立刻就欢迎她来。玛丽回来后，皮埃尔还领她到自己家里。

皮埃尔的父亲是医生。国外医生的待遇是非常高的，是知识分子的上层。法国是当时世界上最民主、最自由、最平等的一个国家。法国人种族歧视很少，特别是在皮埃尔家里面，他父亲曾经是巴黎公社社员，本身就十分同情下层人民。

皮埃尔的父母对于出生于被压迫民族的女孩子一点都没有排斥，很希望儿子和她交往。于是，他们两个在父母的支持下结婚了。他们旅行结婚，两人骑着自行车在巴黎周围的地方转了一下。

默契的合作成就伟大的发现

回来以后他们两个就开始研究铀的放射性。在研究过程当中，居里夫人觉得很可能在铀矿石里面还有放射性比铀强的其他元素，应该把它提炼出来。

居里夫人提议是不是用沥青，从人家修路的沥青渣滓中去提炼。皮埃尔·居里认为可能行，但是这件工作太艰苦了，尤其不适合一个女人去做，于是建议她放弃，但是玛丽坚持要干，最后皮埃尔说好，那就一起干吧。

两个人把学校里一个废旧的、放化学仪器的仓库租下来，那个仓库里面是一个储藏室，外面是个小院。他们就在外面盘了一个灶，支了一口锅熬沥青。仓库里面呢，把实验台清理了一下，架起了一些仪器在那儿做实验。皮埃尔在里面做实验，玛丽拿一个大铁棍在锅里面搅沥青。

每到喝咖啡的时候，两个人就坐在一起商量，下一步应该怎么办。两个人做不同的题目，提炼新元素这件事情主要由玛丽在做，皮埃尔给她出主意，两个人不停地讨论、提炼、实验（图7-1）。



图7-1 居里夫妇在实验室工作

最后终于搞出来了，发现了一种新元素——钋，起这个名字是为了纪念居里夫人已经灭亡了的祖国——波兰。几个月后，又发现了另一种放射性更强的元素——镭。

这些放射性元素的发现立刻轰动了世界。首先，做出这个发现本身就令人震惊，而且是位女科学家做出来的，还是一位被压迫民族的妇女做出来的。居里夫妇和贝克勒尔一起由于发现天然放射性获得了1903年的诺贝尔物理学奖。

在不幸中奋斗

此时，居里夫妇的科研事业正如日中天，可是很不幸，不久皮埃尔·居里就被马车给轧死了，过马路的时候因为思考问题一下被马车给撞了。这件事情对他们的家庭打击非常大，居里夫人一下子失去了事业上的挚友和生活上的伴侣，但是她非常坚强，带着两个女儿，仍然继续做实验，搞研究。

皮埃尔·居里去世以后，他所承担的有关放射性的课没有人能讲，别的男人都不行，只有居里夫人能接手。学校原先不想让居里夫人讲课，因为那时法国大学的讲台只有男人可以站在上面，没有女人讲课的先例。

德国也是那样。德国有位女数学家当时也是讲不了课，希尔伯特后来发火了，说，怎么了？这讲台又不是澡堂子，为什么只有男人能站在上面，女人就不行。

那时妇女走上讲台也是很艰难的一步。法国的大学虽然收女生，但女的主讲教师还没有。由于实在没有其他男人能讲这门课，没有办法，法国这所大学只好破例，让居里夫人上讲台。那天她穿着黑衣服、披着黑纱——丧服，走上讲台讲课。

后来，居里夫人由于创建放射化学，又获得了诺贝尔化学奖。下面我们还会讲到她的大女儿、大女婿得诺贝尔奖。这是令科学界很振奋的一件事情。

西瓜模型和土星模型（图7-2）

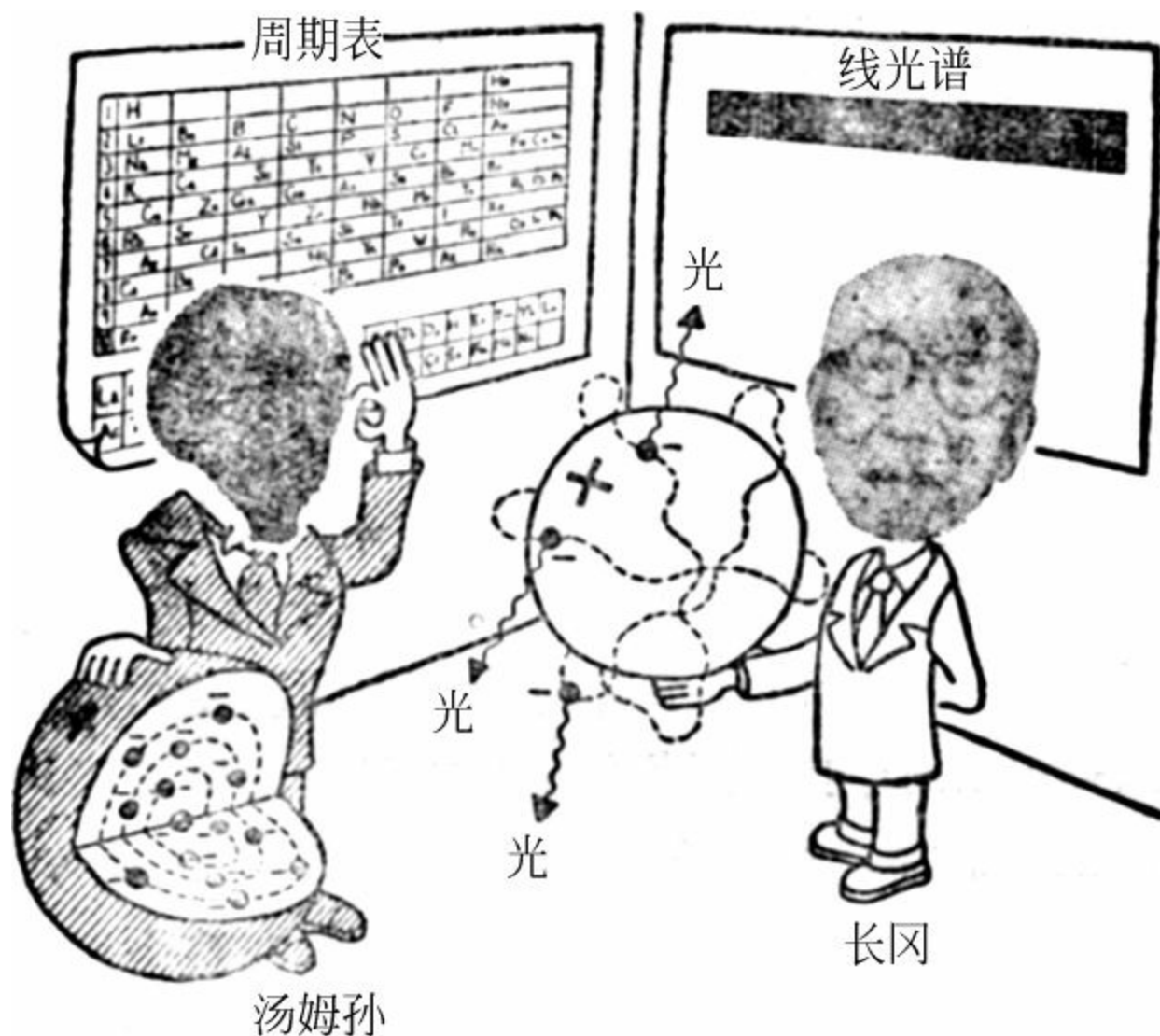


图7-2 西瓜模型和土星模型

1897年汤姆孙发现了电子。原来人们认为原子不可分，后来发现的光谱线、X射线、放射性等现象，都似乎表明原子是有结构的，但又都说不出进一步的东西来。如今发现了电子，原子内部竟然有电子，带负电，那是不是还有一些东西带正电呢。

于是，汤姆孙就提出西瓜模型，这是大家比较熟悉的，就是说，原子就像一个均匀带正电的西瓜一样，其中那些瓜子是带负电的电子，镶

在西瓜里面。这个模型可以解释周期律，但是不能解释光谱线。

日本有一位学者叫长冈半太郎，我们没怎么听说过他，但是卢瑟福的论文当中引用过这个人的工作，他们是知道的。长冈提出了一个土星模型，他也认为，原子是个带正电的球，但是电子并不像汤姆孙的西瓜模型那样位于原子里面，而是在原子外面，像土星的光环一样，有很多电子围着它转。他这个模型能解释光谱线，但不能解释周期律。不过土星模型不同于后来的核模型，长冈认为原子是一个像西瓜一样的实心球，质量和正电荷并不是集中在核心，而是均匀分布在整個球上。

行星模型

后来，卢瑟福发现了 α 射线和 β 射线，不久之后别人又看到了 γ 射线。卢瑟福是新西兰人，是新西兰一个农场主的孩子。他来到英国继续求学，成为一位非常杰出的学者。他发现了 α 和 β 射线，更重要的是他用 α 粒子做了一个散射实验：用 α 粒子去打击原子。卢瑟福是汤姆孙的学生。他想按照导师的这个模型：原子像个大西瓜，均匀带正电的西瓜，那么，如果用带正电的 α 粒子打过去。由于正电荷间的排斥作用， α 粒子打进原子以后会有一个偏转角，卢瑟福经过计算，认为应该是图7-3这样的一个偏转图像。

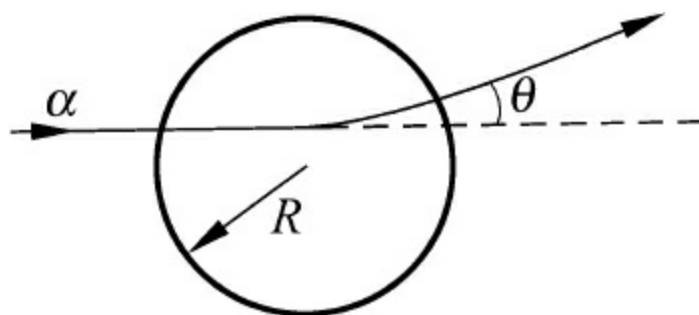


图7-3 西瓜模型预言的 α 粒子偏转

可是实际上，实验中打过去的结果，却是四面八方的散射（图7-4）。只有正电荷集中在核心，才可能产生这样子的散射。由此看来，原子里面基本上是空的，似乎所有带正电的物质都集中在原子的中心，

就是今天所说的原子核。

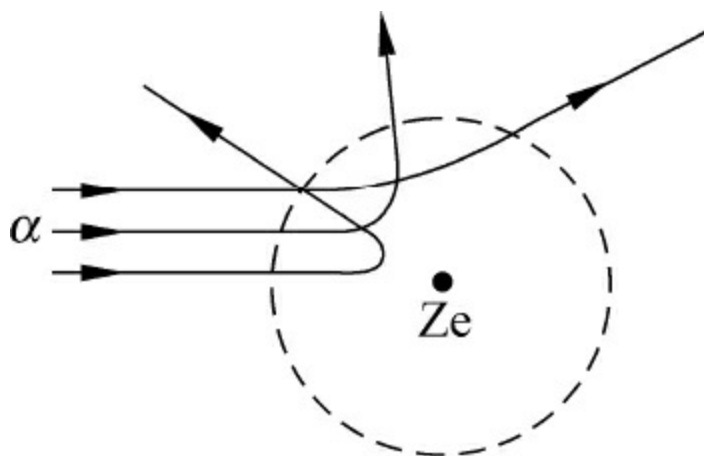


图7-4 α 粒子的散射实验

于是卢瑟福又提出了原子的核模型，中间是个原子核，电子围绕着他转，像行星围绕太阳转一样，这叫做行星模型。

行星模型有一些缺点，它解释周期律和光谱线好像都还有问题。另外，学过电动力学的人都知道，一个带电粒子如果作变加速运动的话，它应该有电磁辐射。电子围绕原子核转，肯定是个变加速运动，它要产生辐射。一旦辐射，能量就会减少，那么电子就会越转圈越小，越转圈越小，最后就会落在核上。因此这个原子模型是不稳定的，这是卢瑟福模型的一个严重缺点。

杰出的卢瑟福

卢瑟福是一位杰出的导师，他培养了很多优秀的学者，培养了11个获得诺贝尔奖的学生，特别著名的就是玻尔和查德威克，还有卡皮查。卡皮查是搞低温物理的。

卡皮查是来自苏联的优秀学生。苏联在十月革命以后，有几次稍微把国门打开一点，派出去一些学者。卡皮查跟一位老先生来到英国考察。考察以后他就不想回去了，但是卢瑟福怕影响英国和苏联的科学交往，不愿意收他当研究生，劝他回去，说“我已经招满了”。“你招多少呀？”“我每年就招30个学生，我已经招满了。”“你就那么严格吗？有点

误差没有啊？”卢瑟福说：“有误差，5%。”“你看，加上我还不到5%”，于是卢瑟福就把他留下来来了。

卡皮查留在卢瑟福那里干得很出色，成为一位杰出的低温物理专家。后来他又去苏联访问，被苏联给扣下来了，让他留下来，为祖国服务。按西方的说法是你干不干，你要不干就把你枪毙了，如果你要干，那么你要多少经费给你多少，给你购置所有需要的仪器。

我觉得这个谈话不太像共产党干部的谈话，他不会说你不干就把你枪毙掉，但可能会强迫你留下来，为祖国服务，你缺什么我们给你买什么。于是，卡皮查就留了下来，苏联也确实买了很多仪器装备给他，从此苏联的超导研究就发展起来了。

卢瑟福后来得了诺贝尔奖。他获得的是化学奖，不是物理奖。当时物理奖和化学奖没有严格的界限。当诺贝尔奖评委会评奖的通知寄来的时候，他的学生从传达室拿到诺贝尔奖评委会的信，“p当p当p”就跑上楼去，“老师啊，诺贝尔奖评委会给你来信了。”卢瑟福和在场的人都很高兴。卢瑟福拿过来一看，就哈哈大笑起来了，说，“你们看哪，他们给我的是化学奖。我一辈子都是研究变化的，不过这次变化太大了，我一下从一个物理学家变成化学家了。”

玻尔模型：轨道量子化

现在我们再来看一下卢瑟福的学生玻尔。玻尔是丹麦人。丹麦古代曾经是一个强国，但那个时候已经比较落后了。丹麦现在在欧洲也不算最发达的国家，有点像欧洲的农村似的，不过它们的农村也是比较发达的农村。因为那个地方的人能吃得了苦，所以丹麦那地方老出优秀运动员。

玻尔对他老师的模型进行了重大改造。大家都知道，就是轨道量子化（图7-5）。玻尔认为，核外的电子只能在若干特定的轨道上运行，这些轨道不会变化。为什么轨道会这样分离，为什么不会变化，玻尔说我不知道，但是只要这样假定，就有可能解释周期律和光谱线，而且原

子结构也会稳定。

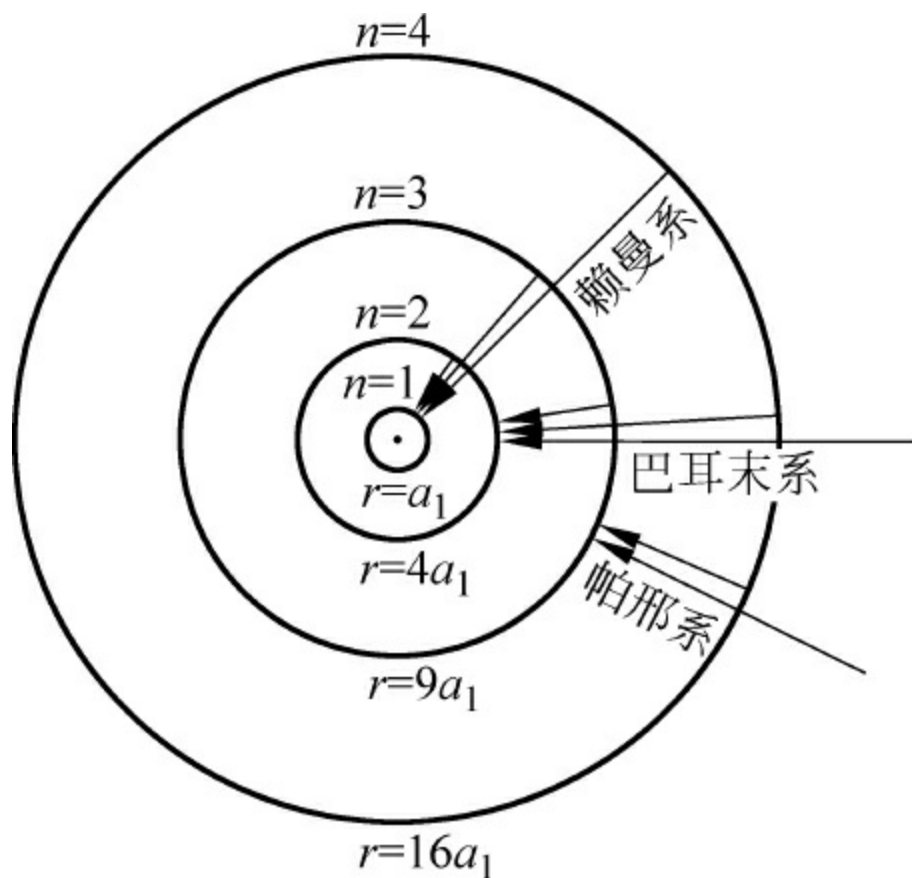


图7-5 玻尔的轨道量子化

泡利不相容原理

当时还有一位年轻物理学家——泡利，他提出来一个不相容原理。因为你有了这个轨道模型以后，人家还会想，既然电子应该往最低的能级跑，为什么电子不都聚集在最下边的能级呢？如果电子都聚集在最下边的能级，光谱线不好解释，周期律也不好解释。

泡利就假定，每一个状态只能容纳一个电子，每个轨道上有两个状态，这叫不相容原理。这两个状态是什么？泡利说不清楚。为什么会有这个原理？也不清楚。有了这个原理，最里层轨道的状态填满后，电子就要填外层轨道，然后再填更外层轨道，一层一层往外排，这样就把周期律和光谱线全都给解释清楚了。当时人们觉得这简直太棒了，把量子

的规律终于搞清楚了。

尖刻的批评家

泡利是德国人，这个人非常聪明。他做实验不行，他到哪个实验室，哪个实验室就出问题，不是瓶子打碎了，就是仪器烧了。但是这个人搞理论特别有一套，他非常聪明。他太聪明了，就总是觉得别人都不行。他预言过中微子，21岁出版过一本相对论。我有这本书的中文版。即使今天再来看，这本书都是高水平的。可他21岁就把这本书写出来了。

泡利曾反对李政道和杨振宁的宇称不守恒猜想。当时李杨两个人说，弱相互作用下左右不是严格对称的，左好像比右强一点，于是就提出宇称不守恒的观念。但是左和右对称多美呀，不对称显然就不太美，泡利跟别人讲，“我就不相信上帝会是个左撇子。听说吴健雄要做实验，我相信吴健雄的实验一定会证明李杨两人是错的。”那时泡利在德国，跟他的同事讲，“你们信不信，我可以和你们打赌，我可以把我全部家产都押上来打赌。”没有人反驳他，也没人和他赌。

过了些日子，吴健雄的实验出来了。泡利在后来回忆时说，“那天下午我一连收到了三封信，都是告诉我说吴健雄的实验支持了李政道和杨振宁的理论，当时我几乎休克过去。现在李杨两个人很高兴，我也很高兴，因为没人跟我打赌，要是有人打赌的话，我就破产了。”

大家知道，杨振宁有三个重要贡献是可以得诺贝尔奖的。有一个比发现“宇称不守恒”更重要的贡献是“杨—米尔斯场”，这是对韦耳规范场论的重要发展。现在所有的规范场理论都是建立在“杨—米尔斯场”基础上的。因为时间的关系，规范场不好讲太多，简单地说，那就是一种物质相互作用的场。

杨振宁搞出这个理论以后，在一次报告会上作介绍。主持会议的是奥本海默（第一颗原子弹的总设计师），泡利坐在旁边。杨振宁刚说了一句，泡利在下边就问了一个问题，说你这个场质量是多少。

什么意思呢？电磁场量子化后是光子，所谓电磁场质量就是光子质量。电子场量子化后就是电子，所谓电子场质量就是电子质量。

泡利就问这个场质量是多少。杨振宁说这个场的质量现在还不太清楚，然后就想继续讲。泡利又说“质量到底是多少”，杨振宁只好面对泡利，说质量现在还不太清楚。泡利说你这理论质量都不清楚，还在这儿讲什么呀。杨振宁当时就没办法了，只好呆站在那里。因为那时杨振宁很年轻，泡利、奥本海默这些人都是老前辈。此时奥本海默就捅泡利，说你先让他讲。于是杨振宁才往下讲了。

杨振宁做完报告的第二天，他住的旅馆房间的外面就放着一封信，一看是泡利的。泡利说像你这种治学态度，我根本就没有办法和你讨论。然后就说你看过谁的论文没有……，据杨先生自己讲，泡利的那个建议还是很不错的。

泡利还批评过那位发现反质子的赛格雷，他是意大利的物理学家。赛格雷有一次做报告，做完报告步出会场的时候，泡利和赛格雷一起往外走，一边走一边说，“你今天这个报告，是我这几年听见的最差的一个。”后面有个年轻人听见泡利讲这个话就笑起来了。泡利一回头看他笑，又说了，“你上次那个报告除外。”

上帝的皮鞭

学者们给泡利起了个外号，叫“上帝的皮鞭”，表示他很厉害，不留情面。为什么叫“上帝的皮鞭”？

“上帝的皮鞭”是对匈奴领袖阿提拉的一个称呼。大家知道，东汉的时候大将军窦宪把北匈奴击败以后，北匈奴越过中亚往西迁移。两百年以后出现在欧洲平原上，把欧洲的那些国家打了个落花流水，欧洲所有的民族都往西移了个位置，给匈奴人腾地方。这叫民族大迁移，欧洲历史上的民族大迁移。后来，公元400多年时又出现一个匈奴领袖阿提拉，率领军队一直打进法国和意大利，欧洲人称他为“上帝的皮鞭”。这就是泡利绰号的由来。

2. 人才特别快车

你学数学没有希望

现在我们该讲量子力学的建立了。先讲一下海森堡。海森堡从小喜欢数学。他的父亲是慕尼黑大学教希腊文的文学教授。海森堡中学毕业以后，想学数学，父亲就去找他们学校的数学家林德曼，想把儿子推荐给林德曼。林德曼让海森堡去面谈。

林德曼据说是研究超越数的，这个方面我不懂。但是林德曼有一项工作，我一说大家就会明白。自古以来几何作图题里面有三大难题。一个是用直尺和圆规能不能三等分一个角，这是大家最熟悉的；还有一个，能不能用直尺和圆规做一个正方形，使它的面积和一个已知圆的面积相等，即化圆为方；第三个问题是能不能用直尺和圆规画出一个立方体，使它的体积是原来立方体体积的两倍，即立方倍积问题。

这三个难题都持续了上千年，无人能解。现在我们知道这三个难题的答案都是不行。而林德曼首先证明了化圆为方不行。

海森堡去见林德曼时，推开他的办公室以后，光线很暗，半天才看清楚，一个白胡子老头坐在桌子后面，抱着一只小狗，那只小狗很不友好，汪汪汪老叫，闹得海森堡心神不定。林德曼问了几个问题，越听林德曼越觉得不行，眉头越皱越紧。后来又问了他一个问题，“你都看过什么数学书呀，看过谁写的数学书呀？”他说看过韦耳的，听完这个话以后，林德曼说了一句，“看来你学数学是没有什么希望了。”

为什么呢？韦耳这个人搞的不是纯数学，是应用数学，大家知道韦耳是非常杰出的数学家。他在物理上的贡献是很大的，但在纯数学上可能贡献不大，他搞的不是那种纯数学。

另外大概还有其他的问题。哥丁根大学数学界的内部有一些矛盾。其中一个矛盾是，当时他们学校里有一位非常漂亮的女士，好几个年轻的数学才子都追求这位女士，最后被韦耳追到了。大概其他人，觉得韦

耳追到了，有点遗憾还是怎么着，搞不大清楚。比如说冯·卡门，钱学森的那位老师也是失败者之一，也没追上。

索末菲的人才快车

海森堡一看反正学不成数学了，他想那就学物理吧。于是他父亲就又介绍他去见物理教授索末菲。海森堡去见索末菲，索末菲模样很威严，穿得非常整齐，西装革履，胡子翘着，很像普鲁士军官，但是说话很和蔼。问他，“你看过什么书，你喜欢什么。”听他讲了以后，说好吧，我收下你。但是你要注意，作为一个初入门的学生，首先你要立大志做大事；另外要从简单的问题做起，要先易后难，积累经验，树立信心，然后去做更难题目；并鼓励他要勤奋地做练习。

索末菲培养学生有一套办法，他办了一个由研究生和优秀本科生组成的研讨班，叫做“人才特别快车”，其中有泡利，还有海森堡，泡利先在这个班学习，毕业后留在这个班当助教。

学术的天堂：哥丁根大学

除去慕尼黑大学之外，对海森堡产生重大影响的还有哥丁根大学。哥丁根大学是全世界第一所在教学和科研上具有充分自由空气的大学。那所学校所在的小城整个就是一座大学城，里面的居民大都是为这个学校服务的，比如说给学生提供宿舍、午餐、……

这所大学学术氛围非常之浓厚。教授讲完课以后，走在半路上碰到学生，学生会在路上问老师问题，最后学生越聚越多，在马路上围成了一圈，听那位教授在街上演讲。学生们讨论问题的时候也是如此，有时候白天习题没有做完，半夜里有人做出来了，就跑出去找他的同学，咣咣敲窗户，喊这题我做出来了。还有个学生，摔了个跟斗，一下倒在地上，别人要来扶，他说“别扶，别扶，别打扰，我的问题正有点儿开窍呢”。

这所学校是非常值得注意的，你们可以看一看有关资料。有一本书叫《比一千个太阳还亮》。这本书是讲美国制造第一颗原子弹的，但其

中有一大段内容是介绍哥丁根大学的。

我上大学的时候看过那本书，看了讲哥丁根大学的那段以后，内心震动非常之大。我觉得我们办大学的方法有些是不对的，是有很严重缺陷的。也就是说，不要用办中小学校的办法来办大学，大学应该是一个自由讨论的，具有自由学术空气的地方。

大家来看哥丁根大学，这个学校在数学方面，有高斯、黎曼、克莱因、希尔伯特、韦耳、冯·卡门、冯·诺依曼等，这些数学家都是在那里工作或学习过的。物理方面有玻恩、劳埃、奥本海默、康普顿、狄拉克、鲍林、洪德和约丹。这一个个物理学家也都是从那里出来或者在那里工作过的。所以那所学校真是不简单啊，真是我们学习的榜样。

愚蠢的问题受欢迎

当时玻恩在哥丁根大学主持一个理论物理研讨班，数学大师希尔伯特经常来听。这个研讨班有一句格言叫“愚蠢的问题不仅允许，而且受欢迎”，就是要自由讨论，鼓励青年人勇敢发言。因为数学逻辑是很严谨的，而搞物理的这帮人呢，是一边猜想着一边研究。希尔伯特听完他们的讨论之后觉得：哎呀，怎么这个样子啊！希尔伯特感叹之后就说了句话：“看来物理学对于物理学家来说实在是太困难了！”讽刺了一通。

由于处在世界数学研究的中心，又经常有数学家光临，玻恩的课题组，比其他大学的物理课题组更加偏爱数学。海森堡刚开始到那里觉得很是不习惯。海森堡是在慕尼黑大学学习和工作的，后来，索末菲介绍他和泡利到哥丁根大学去看一看，这样他们跟哥丁根大学就有了一些来往。

玻尔曾到哥丁根大学访问了十天，这十天在哥丁根大学的历史上叫玻尔节。索末菲带着泡利和海森堡去参加。海森堡和泡利在会上向玻尔提问题，索末菲又带着他们和玻尔一起出去散步，玻尔对这两个年轻人非常器重，就对他们说：“欢迎你们随时到哥本哈根来！”然后玻尔就回

国了。

3. 矩阵力学与波动力学

当时量子力学有3个带头人，一个是索末菲，他比较偏重实验。一个是玻恩，比较偏重数学，再一个就是玻尔，比较强调物理思想，因为时间关系我就不讲那么多了。我在咱们学校有一个30课时的讲座，叫“从爱因斯坦到霍金的宇宙”，时间比较充分，我会讲得更详细一些，你们要是愿意，可以去听。

海岛上的灵感：海森堡创建矩阵力学

现在来讲真正的量子力学的建立。真正的量子力学，首先是从矩阵力学开始建立的。创建矩阵力学，海森堡的贡献最大。海森堡在玻恩的启发下，认识到看不见的东西其实并不重要，重要的东西是我们能够看见的，实验上能够测到的东西。他注意到玻尔说的那个电子轨道，谁也没有看见过。能看见的是什么呢，是光谱线，光谱线反映的是两个轨道的能级差，而不是轨道本身。所以海森堡认为：重要的不是轨道，而是轨道之间的能量之差。能级差，才是最重要的。

1925年的春夏之交，海森堡患上一种过敏性疾病，在北海的一个小岛上疗养。在轻松悠闲的生活中，他的思想终于有了飞跃。他创造了一套符号，用这套符号，不依赖玻尔模型就能算出光谱线来。玻尔他们知道后很感兴趣，也都很高兴。

玻尔这个人非常大度，跟学生总是平等地讨论，从来不摆架子。所以他能吸引一大批年轻人在他的周围，比如说海森堡、泡利、狄拉克、朗道这些人都在他身边工作过。

海森堡成功地创造了一套符号，但这套符号不满足乘法的交换律。海森堡当时很担心，觉得“这东西怎么会不满足乘法交换律呢，我这里面说不定会有问题，如果有问题就会在这个地方！”其实他的理论正确就正确在它不满足乘法的交换律上，正是不满足乘法交换律，把量子效

应引了进来。

这一年海森堡写了一篇文章，海森堡的数学水平不是特别突出，但他的物理思想非常棒，他写的文章物理思想很创新，但文章结构不是那么严密。文章出来以后玻恩看见了，觉得：“海森堡文章中的这些符号有点像矩阵啊！”因为玻恩在哥丁根大学，哥丁根是数学中心，所以他知道数学中有矩阵这种分支。玻恩这么一讲，他身边的一个年轻人约丹就说：“玻恩教授，我熟悉矩阵，咱们俩一块儿干？”玻恩同意了，于是两个人一起在海森堡工作的基础上，完成了一篇文章。然后他们又把海森堡拉进来，三个人又搞了一篇文章，这在历史上就叫做“一个人的文章，两个人的文章和三个人的文章”，这三篇文章就把矩阵力学建立起来了。玻尔他们非常高兴，觉得把轨道模型的理论往前推进了一大步。

山峰上的灵感：薛定谔创建波动力学

正在他们高兴的时候，却听说瑞士那边有一个叫薛定谔的物理学家，搞了一套波动力学，没用他们的矩阵，只用微分方程就把那些光谱线也算出来了，他们觉得很奇怪。这个时候索末菲就邀请薛定谔到慕尼黑大学做报告。

薛定谔建立波动力学的过程是这样的。先是法国的德布罗意提出了物质波的理论，认为正像光波具有粒子性一样，电子等实物粒子也具有波动性。

公式

$$E = h\nu \quad (7.1)$$

和

$$p = \hbar k \quad (7.2)$$

就是德布罗意波的表达式，其中E、p分别为粒子的能量和动量， ν 、k分别为对应的德布罗意波的频率和波矢，其中

$$k = \frac{2\pi}{\lambda} \quad (7.3)$$

$$h = \frac{h}{2\pi} \quad (7.4)$$

λ 为波长， h 为普朗克常数。

德布罗意的老师郎之万不知道这个理论对不对，就把有关的东西寄给了爱因斯坦，爱因斯坦看后非常赞赏。

奥地利物理学家薛定谔当时在瑞士工作，他们那个学校周末的时候经常有一些学术报告，有一次薛定谔在做完报告以后，主持会议的教授德拜对薛定谔说：“你今天讲的那些没有太大意思，听说德布罗意搞了一个物质波，你能不能下次把他的这个工作介绍介绍。”薛定谔说：“可以！”在第二次会议的时候薛定谔就讲，讲完以后德拜就问了：“它既然是个波，它的波动方程是什么样啊？”薛定谔想，对啊，德布罗意波没有方程啊！于是他就去努力寻找。过了一段时间以后，他又出来做报告说：“上次德拜教授说德布罗意波没有方程，我现在找到了一个！”

这就是著名的薛定谔方程，量子力学最基本的方程。他是用经典力学和光学的类比建立起来的，他有一本小册子叫《关于波动力学的四次演讲》，你们可以翻一翻。

公式

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2\mu} \Delta^2 \psi + V\psi \quad (7.5)$$

就是薛定谔方程。式中 μ 、 V 分别为粒子质量和势能， ψ 为对应的德布罗意波的波函数。

薛定谔是在1925年末和1926年初的时候把这个方程找到的，当时他跟他的一个女友在阿尔卑斯山上度圣诞节和新年。薛定谔年轻时候曾经有一个非常喜欢的姑娘，但是被这个姑娘的母亲给坚决切断了。这位母亲觉得薛定谔是平民出身，配不上自己贵族家庭的女儿。这件事对薛定

谔打击很大，但也可能起了积极的反作用。有人认为，当初薛定谔如果跟这位最喜欢的姑娘结婚的话，也许他就很满足了，他可能就做不出什么发现了。

这次在阿尔卑斯山上跟他共度圣诞节的女友是谁？不清楚。有人认为，这位女友与他的共同生活激发了他的灵感，对他建立波动力学起了积极作用。所以有一些传记作家，专门到那里去调研和考察，最后也没考证出来这位女士到底是谁。

刚开始的时候，薛定谔建立的是一个相对论性的波动方程，但这个相对论性的方程跟实验对不上，他后来又退回到非相对论情况，才把那个正确的方程得到。为什么相对论的不行，非相对论的反而行呢？相对论性的方程是跟粒子的自旋有关的，不同自旋的粒子有不同的波动方程。薛定谔得到的那个方程对于自旋为零的粒子是对的，但是电子自旋是二分之一，所以这个方程不行。非相对论的波动方程跟自旋无关，因此他退到非相对论情况反而成功了。

海森堡与薛定谔的初次交锋

薛定谔方程出来以后轰动非常大，因为物理学家都熟悉微分方程而不知道矩阵。今天在你们看来，矩阵其实很简单，可是当时的物理学家都没听说过这个东西，所以海森堡的影响，他们那套矩阵力学的影响还不如薛定谔的波动力学大。薛定谔的波动力学一出来大家就都注意到了，于是索末菲邀请薛定谔到慕尼黑大学来报告一下他的工作。

1926年7月薛定谔来了。他作报告的时候，教室挤得满满的。他讲完以后，坐在听众当中的海森堡就从拥挤的人群里站起来，提了几个问题，说：“你这个理论里没有不连续性，它怎么能得出量子效应呢？”然后又问了一些问题，把薛定谔给问住了，弄得薛定谔很难堪。这个时候主持会议的维恩站起来示意海森堡坐下。说：“薛定谔教授的理论当中的问题，他自己会慢慢解决，你还是把你的功课弄得好一点！”

维恩怎么这么说话呢？原因是此前不久，他参加了海森堡的博士学

位答辩，海森堡的导师是索末菲。答辩委员会主席就是维恩，维恩是位实验物理学家，海森堡的论文是搞流体力学的，本来他这论文讲的也没有什么问题，好像就可以通过了，结果在无意之中维恩问了一个关于光学误差的问题，海森堡不会，这就引起维恩注意了。又问他说“你跟我讲一讲法布里—珀罗干涉仪是怎么回事啊？”他讲不出来，维恩一看“嗯？怎么讲不出来呢？”又问“你跟我讲一讲显微镜的原理吧。”海森堡也不会，然后说“那望远镜呢？”他也不会。“那蓄电池呢？”还不会。维恩想，这个家伙简直是连本科毕业的水平都达不到，怎么能给他博士学位呢，就不想让海森堡通过。在他们讨论的时候，海森堡的导师索末菲出来力保。最后答辩委员会一投票，索末菲给了一等，还有一位也给了一等，另外的人呢给了二等，那个维恩坚持给了海森堡一个四等，就是不及格。结果折中了一下，给了他三等，也就是将将及格。当时海森堡很恼火，本来博士答辩完了之后有一个酒会，他连酒会都没参加就跑了，当天晚上直接去了哥丁根大学，向玻恩诉苦去了。玻恩还不错，说没关系，成败不决定于一两件事，还是欢迎海森堡经常到他们那里去。

在薛定谔的报告会上，维恩还记着这事，又冲海森堡来了，海森堡当时很不高兴，第二天就不参加这个会了，出去玩去了。他同时给玻尔写了一封信，告诉他这边讨论的情况。

4. 关于量子力学本质的大论战

我真后悔来这里：玻尔与薛定谔的论战

玻尔也知道薛定谔这个工作还是很重要的，于是邀请薛定谔到他那里访问，薛定谔到了哥本哈根，玻尔到车站去接他，寒暄了几句之后玻尔就开始问问题，之后一直到开会，继续问问题。玻尔手下是一批非常优秀的年轻人，问题提得非常尖锐，他们问“你这个波是个什么波啊？”“粒子和这个波是什么关系啊？”薛定谔说“粒子就是‘波包’”，但是那些人反应很快，说“在真空当中，德布罗意波各个频率成分的波速

是不一样的，要发生色散。要是基本粒子是‘波包’的话，这‘波包’一会儿就散开了，消失了。”问得薛定谔面红耳赤答不出来。

两天以后薛定谔就病倒了。玻尔去看他。薛定谔说：“我真是后悔，我干吗要来这儿呢！”玻尔说：“你不能这么想，大家提问题是因为你的工作不错，假如你的工作不行，我们就不会问那么多问题了。”然后玻尔又继续问他问题。

薛定谔呆了几天以后，很不愉快地离开了哥本哈根。回去以后他又进行了一些研究，1926年的下半年他证明了波动力学和矩阵力学是等价的，以后就把矩阵力学和波动力学合在一起称为量子力学。

论战愈演愈烈：概率波与测不准关系

1926年，玻恩提出了波函数的概率解释，认为德布罗意波是概率波。波函数的模的平方表示粒子出现的概率。

后来，海森堡又在1927年提出了测不准关系，如下两式所示，

$$\Delta x \Delta p \sim h \quad (7.6)$$

$$\Delta t \Delta E \sim h \quad (7.7)$$

式（7.6）表示位置和动量不能同时确定，这样粒子就没有轨道了。式（7.7）表示时间和能量也不能同时确定。例如激发态的能级宽度和电子在这个能级上呆的平均寿命之间，满足这个测不准关系，能级越宽越不稳定，电子呆的寿命越短。第二个测不准关系一般人不大注意，一般比较注意第一个。

量子力学诞生之后，一直伴随着论战，它的主流派称为哥本哈根学派，承认测不准关系。这个主流派的理论是在矩阵力学的基础上建立起来的，他们也认识到了矩阵力学和波动力学是一致的。这个学派认为波动力学中的波是一种概率波，表示粒子在一点出现的概率。量子力学只能告诉我们粒子处于某种状态的概率，不能确定地告诉我们是否一定在某种状态出现。

他们这个理论遭到了一些很优秀的物理学家的反对，比如爱因斯坦、薛定谔和德布罗意，还有后来的玻姆，他们坚决不同意这个观点。双方经常论战，反对的一方不断地提出各种反例来说明概率波的理论不对，但是都被哥本哈根学派的人一个一个地驳倒。

薛定谔始终不愿意承认概率波这套理论。爱因斯坦也认为概率的描述肯定不是最终的理论。有一些玻尔那边的年轻人就讽刺薛定谔，说“看来薛定谔方程比薛定谔本人更聪明”。

薛定谔引领生物学革命

是不是这样呢，历史表明，不是这样。薛定谔非常了不起。薛定谔是个大器晚成的人。当时的许多杰出学者都是二十几岁就做出成就了，薛定谔是三十九岁才创建波动力学的。

薛定谔后来又创出新成就，在生命科学方面作出了划时代的贡献。他1943年在爱尔兰都柏林的三一学院发表了连续演讲，题目是《生命是什么？》。

薛定谔提出几个重要观点，一个是“生命来自负熵”，这个观点很重要。一般人都以为人们吃东西只是为了补充能量。维持生命最重要的条件是补充能量，能量是生命的源泉。薛定谔指出这种看法不对，没有抓住本质。如果只是为了补充能量的话，咱们都不用生产粮食了，只要挖煤就行了，只要把这屋里的暖气烧得比 37°C 高一点，那热量不就往身体里面流吗？但是经验告诉我们，这样并不能维持生命。

薛定谔指出，关键在于生命需要负熵来维持。需要高质量的能量，低熵的能量。与低熵相伴的能量才是生命的可用能，或者叫有用能。与高熵相伴的热能是比较低级的能量，对生命用处不大。他认为，生命的关键不在于能量，而在于负熵。这是一个非常重要的结论。

他又说，“遗传密码的信息存在于非周期的有机大分子当中”，历史表明他的预言是正确的。现在的DNA理论和他的预言正好相符。

他还指出，“生命是以量子规律为基础的，量子跃迁可以引起基因

的突变”。这一预言也已被实验所证实。

DNA的双螺旋结构的两个发现者——生物学家沃森和物理学家克里克，青年时代都读过薛定谔的《生命是什么？》这本重要著作。这本书有中文版，你们可以找来读一读，一定会有所收获。

单电子引来的疑难

我们现在要对波函数的概率解释说几句。先来看一下经典粒子的运动（图7-6）。比如说有一支枪往靶上打子弹，枪和靶之间有一个屏，上面有两个缝，把下面的缝2关上，只开上面那个缝1的话打出来的强度（即打在靶上的子弹数密度）分布是 P_1 这根曲线，要是把1这个缝关住开2的话打出来的强度分布是 P_2 这根曲线，你要是两个缝都打开的话打出的是 P_{12} 这根曲线，它正好是 P_1 与 P_2 两根曲线的叠加。这是粒子的情况。

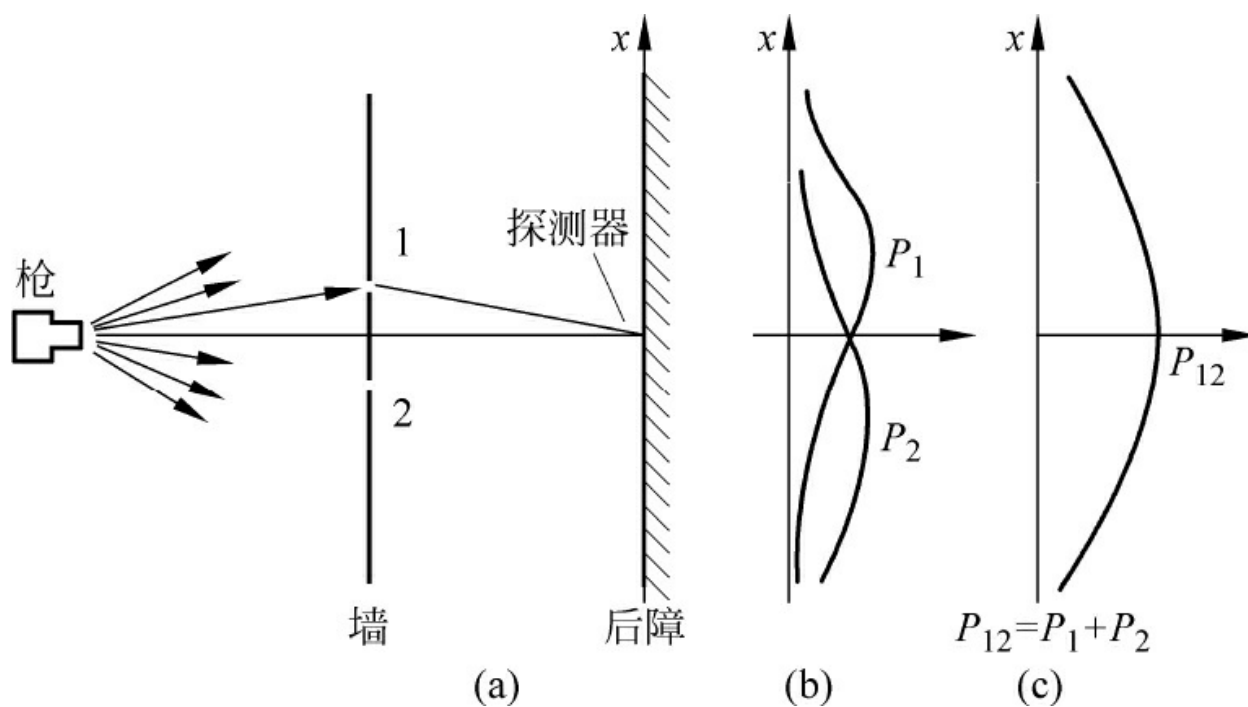


图7-6 子弹双缝实验

那么波的情况呢（图7-7），你把缝1挡住，它的强度分布曲线是 I_2 这根曲线，如果你要把2这个缝挡住呢，得到的是 I_1 这根曲线。两个缝

都打开，就会有干涉现象，出现的强度分布曲线不是 I_1 和 I_2 的简单叠加，而是后面的有干涉条纹的这条线。

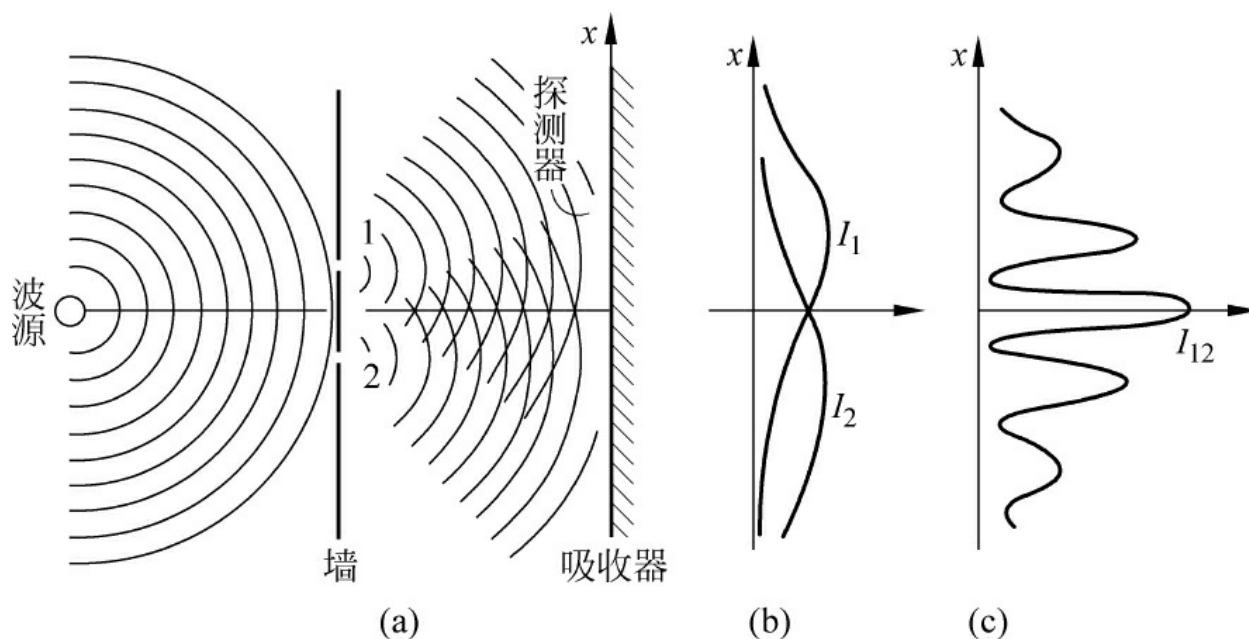


图7-7 水波双缝实验

现在用电子来打靶（图7-8），这么小的粒子打靶的话，就会发现，跟波的情况非常相像。你把2这个缝挡住它打出来的强度分布是 P_1 ，你要把1挡住，打出来的是 P_2 ，两个都打开，居然出现了干涉条纹，这就说明了电子是波，电子的运动具有波动性。

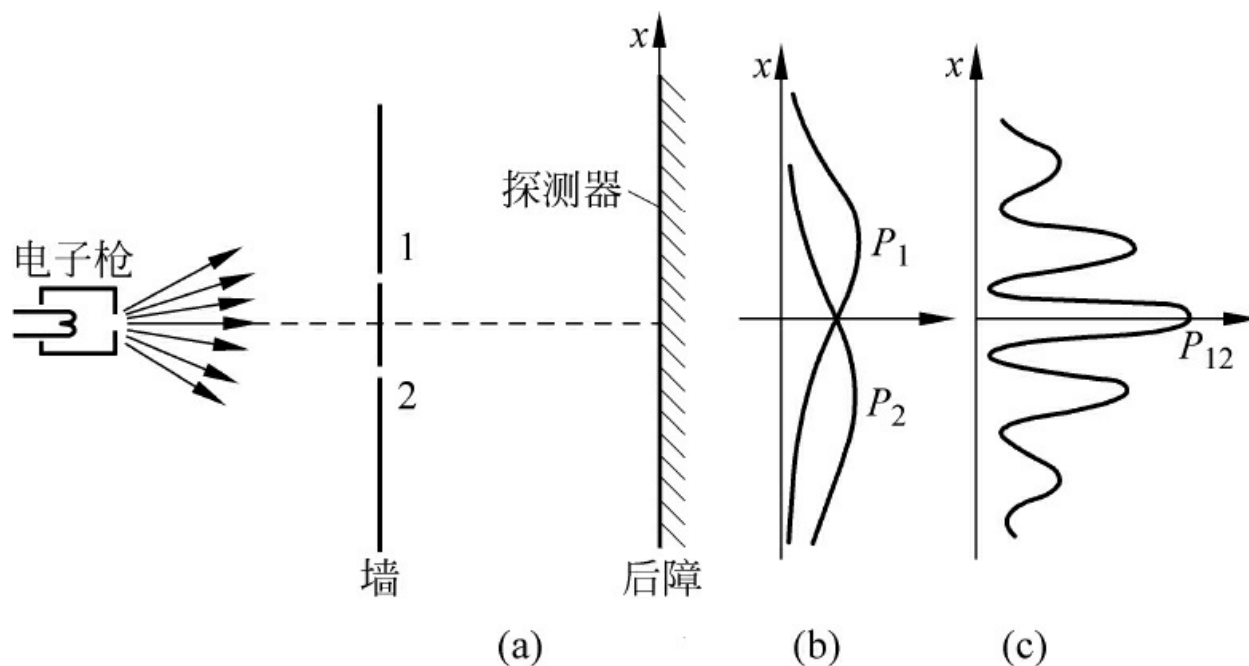


图7-8 电子双缝实验

有人猜想这或许是电子之间的相互作用造成的，是大量电子的行为，如果电子一个一个发射，可能就不会出现干涉条纹了。

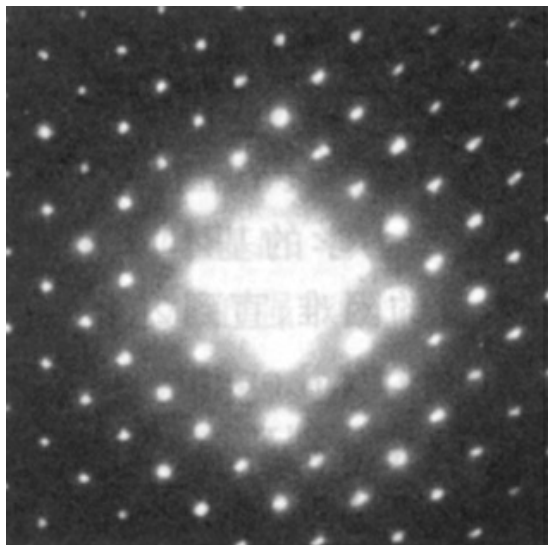
后来人们做了单电子干涉实验，就是让电子枪里边出来的电子几乎是一颗一颗的，它们之间相互没有影响，结果仍然打出了干涉条纹。这个问题，一直到现在，都有很多人觉得不清楚。就好像一个电子能够同时穿过缝1和缝2，其行为就像图7-9所画的这个量子滑雪者一样，相当诡异。甚至有人猜测是不是电子有“自由意识”，在过缝1的时候，它知道缝2开没开。因为电子波如果是概率波的话，电子作为一个粒子似乎它只能走缝1或者只能走缝2，似乎不可能同时走缝1和缝2。但这又如何理解干涉条纹的出现呢？只好想象它走缝1的时候知道缝2开没开，它走缝2的时候又知道缝1开没开，只有这样才会出现干涉条纹。这太令人难以理解了，所以这个问题一直引起大家很大的兴趣。



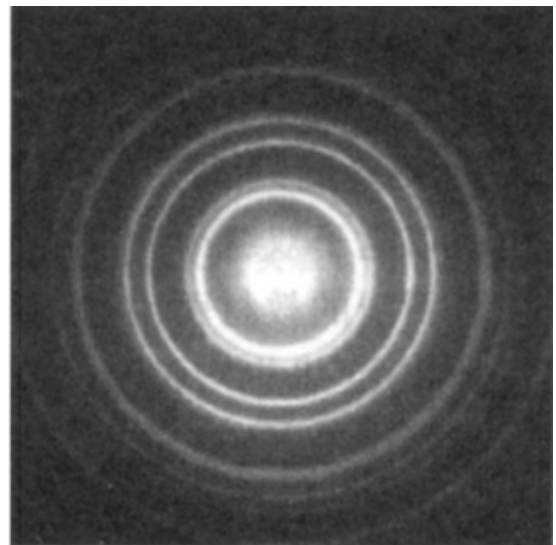
图7-9 量子滑雪者

概率波实验支持

大家知道，其实我们打出来的干涉条纹，我们通常看到的干涉条纹，是大量的光子或者电子形成的条纹（图7-10）。如果粒子（光子或电子）是一个个打过来的话（图7-11），刚开始屏幕上没有点，或者有一个点，然后稍微多一些点，点打得多了，就逐渐地形成了条纹。



(a) MoO_3 单晶的劳埃相



(b) Au多晶的德拜相

图7-10 电子衍射的劳埃相和德拜相

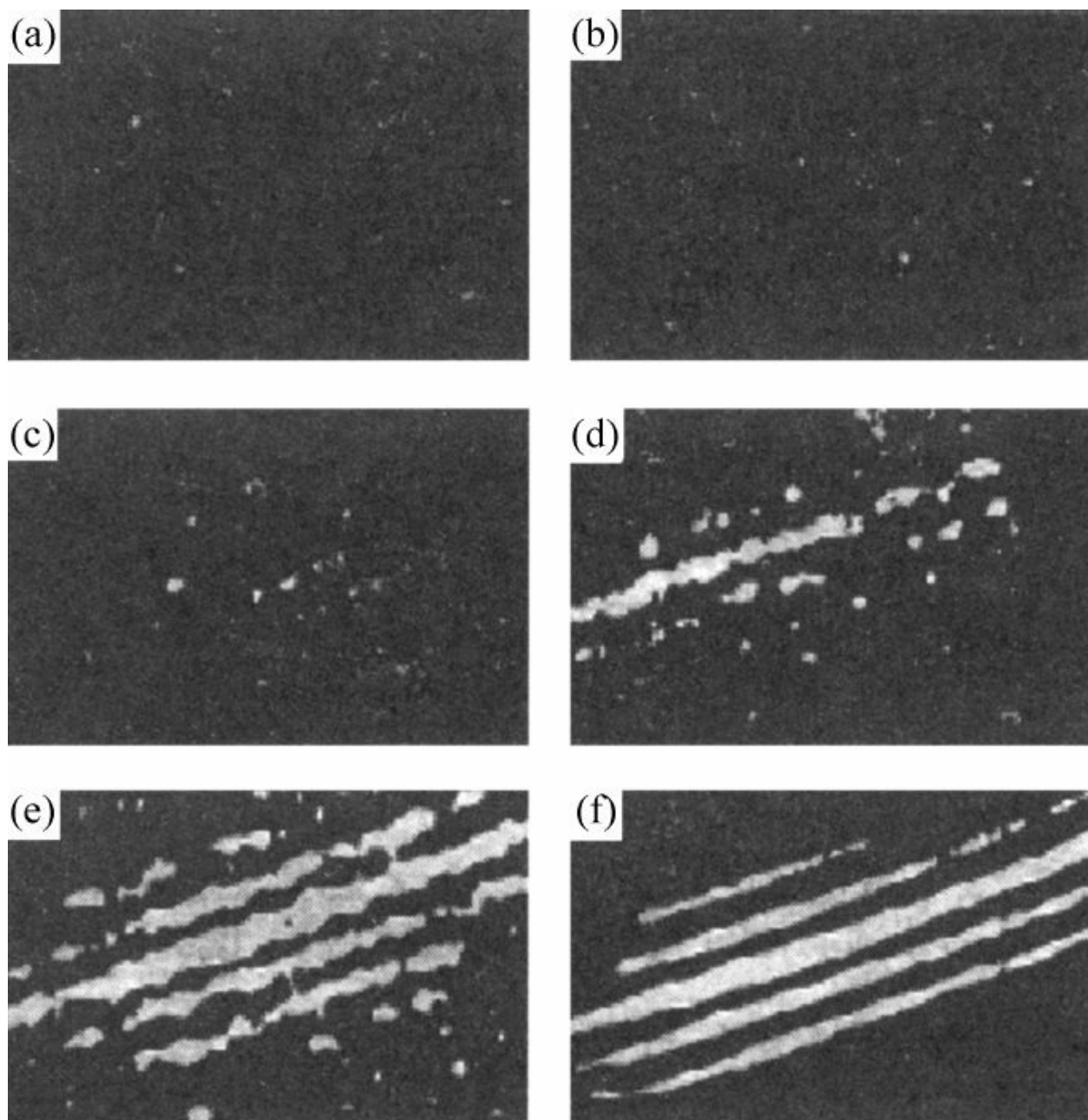


图7-11 真实实验中获得的电子干涉图样

也就是说，这个波，粒子对应的波，它传播的过程是以波动的形式，但是相互作用的时候是粒子的形式，集中在一个点上，打得多了，才逐渐形成干涉条纹。这种波为什么会是这个样子？当然用量子力学是可以严格算出来的，但是许多人还是觉得在物理图像上很难让人理解。我想一般人都会觉得确实很奇怪。

神奇的隧道效应

还有一个有趣的问题是隧道效应（图7-12）。有一个高的势垒，一个经典粒子射过来，如果粒子的能量低于这个势垒的话，粒子就肯定过不去，如果能量高的话就肯定过去了。



图7-12 隧道效应

但是量子的情况却不同，粒子过来的时候如果它能量比势垒低，仍会有一定的概率穿过去，当然也会有一定的概率被反射回去。如果能量比势垒高，也会有一部分被反射，不会都越过去。这就是量子力学中的隧道效应。

为什么会有隧道效应，粒子能量不够为什么也会越过去？现在一般的解释是：这个粒子过势垒的时候借用了能量，利用测不准关系提供的可能性，它从“虚无”中借用了能量。越过去后再把借用的能量还给“虚无”。由于测不准关系的限制，粒子借用的能量 ΔE 越多，它可借用的时间 Δt 就越短。因此粒子穿越势垒的过程必须很快。有时甚至必须超光速。这是一种直观的理解。

有一种观点认为，穿透势垒好像是瞬时的，好像不需要时间，为什么会是这样子还不太清楚。穿过势垒不需要时间，这件事情有些人研究过。

我们最近在弯曲时空当中也研究过这个问题。在平直时空中，假如粒子穿过势垒不需要时间的话，它进入势垒的时刻 $T=1$ ，那么从势垒穿出的时刻也是 $T=1$ 。可是在有的弯曲时空当中，两个时间点“同时”，并不意味着两个点的时间值等同，它会有一个差，“同时”的时刻值由于时

空弯曲会不相等，会有一个有限的差值。我们就按照那个差值来算，得出来的是跟其他理论相符的结果。

我们在一篇计算黑洞辐射概率的论文中，研究了黑洞附近的势垒贯穿效应，结果发现粒子“瞬时”穿过势垒的观点是有道理的。

暗箱中的粒子

再看一个粒子在箱中位置的实验（图7-13）。有一个暗箱，隔成A和B两个部分，隔板上有小孔，电子可以通过。箱子里头有个电子。它在A还是在B？大家想，它不是在A就是在B。

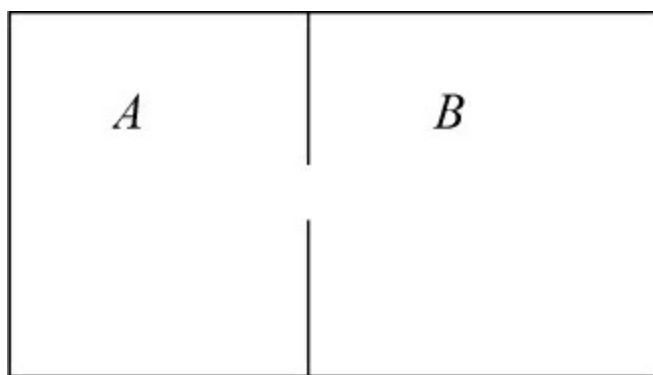


图7-13 暗箱中的粒子

但是量子力学计算的结果是，你打开这个箱子看之前，电子同时在A和B，它是同时处在这两个部分，那打开箱子呢，你一开箱它就缩到A了，或者缩到B了。但你开箱之前，它不是说已经在A或者已经在B了，你一开，一观察，它就缩到一个点上去了。

很多人觉得这好像让人不可思议。也就是说，在观察之前，电子是处在状态1和状态2这两个波函数的叠加态中，电子的波函数是一个叠加的波函数。你一观察，它就缩到状态1或者缩到状态2了，它不会在观察时还同时处在状态1和状态2两个状态中。

既死又活的猫

许多人觉得哥本哈根学派的上述理论很难让人接受，薛定谔就提过一个反例，叫薛定谔猫（图7-14）。他说，把一只可怜猫关在一个箱

子里，这个箱子内部上面有个电磁铁，是个继电器，吸着个铁锤，底下有个装氰化氢毒药的瓶子，旁边有个放射性原子，这个原子如果衰变了，发射粒子了，打进这个盖革计数器，计数器一接收粒子，继电器就断电了，一断电，铁锤就掉下来，把毒药瓶打碎，猫就死了。如果这个原子在这段时间里头没有衰变，它不发射粒子，继电器就不断电，毒药瓶就不会碎，这猫就活着。

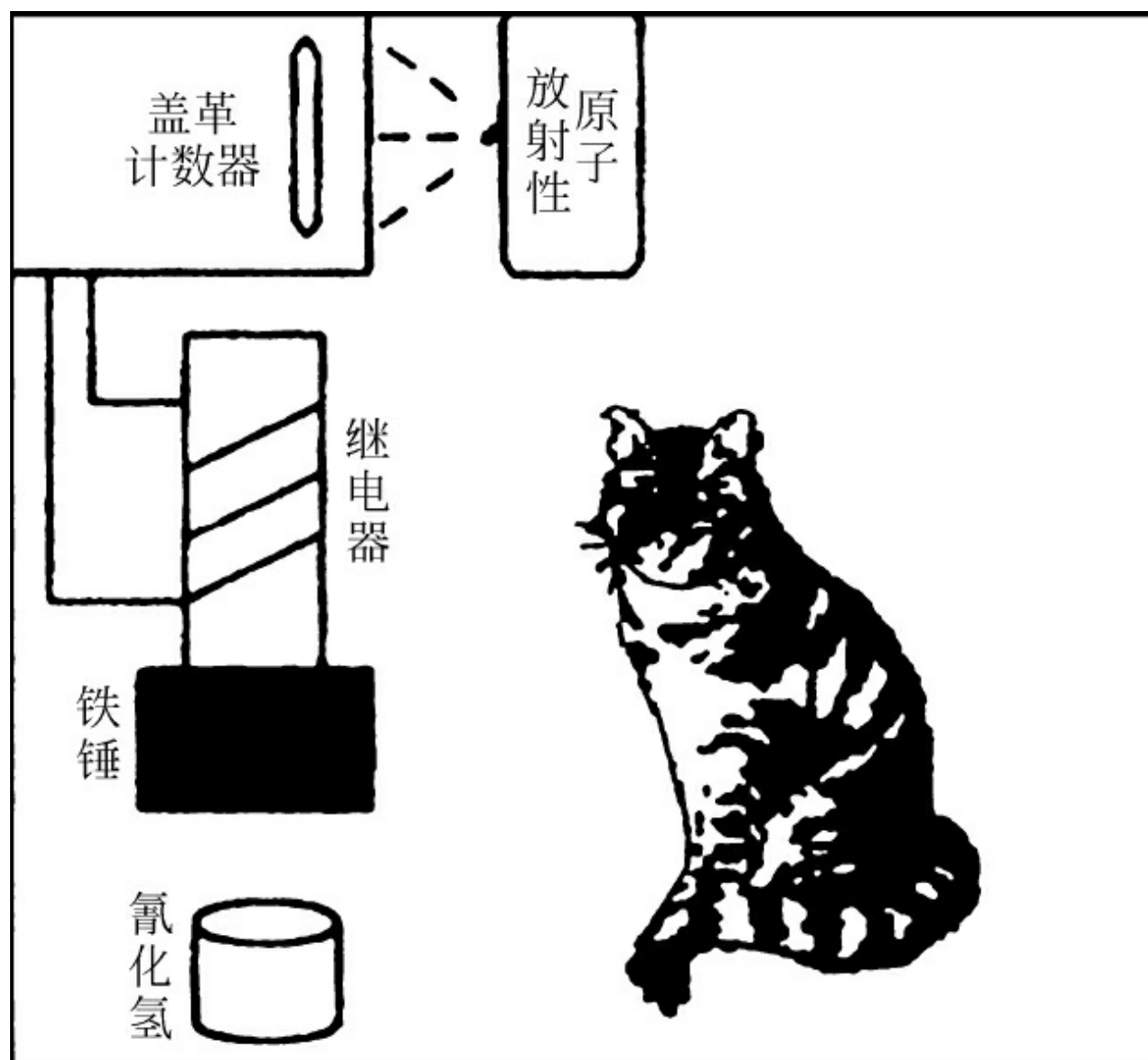


图7-14 薛定谔猫

这样就会有一个问题，在你开这个箱子之前，这猫处于什么状

态，是死还是活？你打开箱子，它肯定不是活着就是死了。那么开箱之前是什么样，根据量子力学的理论，开箱之前它处在既死又活的状态。而且按“既死又活的状态”计算的结果，跟开箱之前它就已经死了或者依然活着，这样计算出来的结果是不一样的。也就是说，如果你承认量子力学的话，你就真地承认开箱前，猫同时处在既死又活的状态，但是你一开箱它就死了或者活着。

爱因斯坦、薛定谔一方觉得这个思想实验表明量子力学的统计解释有问题，玻尔一方也无法解释清楚。爱因斯坦讽刺说，我就不信，一只老鼠，你看它一眼整个宇宙都会变化了，这怎么可能呢。这个例子一直到现在还在争论，学术界还在争论薛定谔猫这个反例，当然整个的讨论还是有利于哥本哈根学派的。

爱因斯坦的光子箱

1930年，在第六次索尔维会议上，爱因斯坦向玻尔提出挑战说，你看我能同时确定能量和时间。你们那个测不准关系，不是说能量和时间不能同时确定吗？你看，这儿有个箱子（图7-15），箱子上面有个弹簧秤，箱子里头有个钟，箱子旁边有个小口。

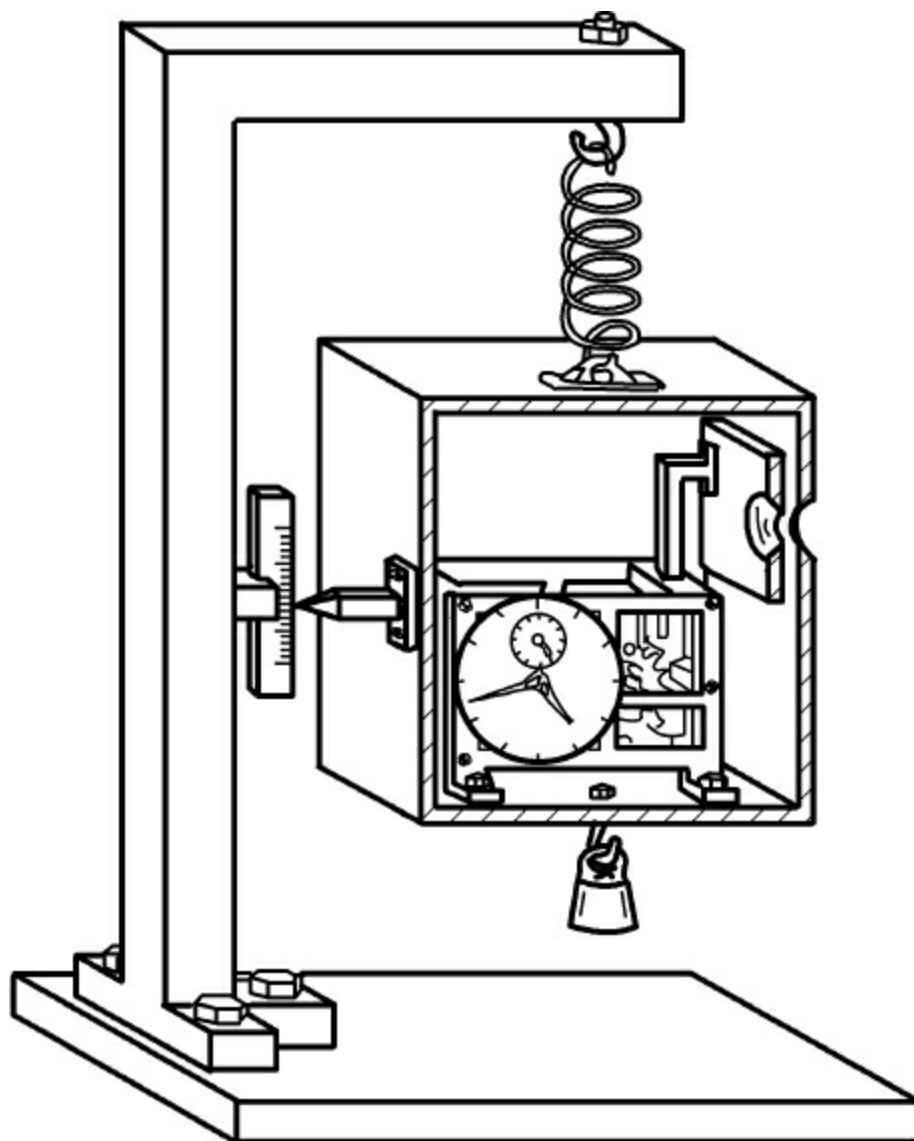


图7-15 爱因斯坦的光子箱

爱因斯坦说，设想一个光子从这个小口出去。我打开这个口，噫一下光子出去了，它的质量不就减少了吗，减少了一个光子的质量，那么弹簧秤指针就会移动，所以我就可以从弹簧秤指针位置变化知道出去的这个光子的能量，又可以通过钟表指针知道它出去的时刻，这两项测量互不影响，这不就把光子射出的时刻和光子能量两个量同时精确确定了吗，怎么会遵从你那个时间能量测不准关系呢。

提出这个反例以后，当天晚上有人对玻尔说，爱因斯坦肯定又是错

的。但是玻尔觉得很震惊，觉得这个反例还真的不好回答。他一晚上没睡好觉，在那儿研究这个问题。玻尔在半夜里经过仔细思考，最后得到一个答案。

这个答案总的意思是说，光子射出时盒子会获得一个向上的动量 p ，它来源于光子逸出造成的冲量。然而盒子的指针位置 x 会有不确定度 Δx ，它来源于盒子动量的不确定量 Δp ，及位置动量不确定关系 $\Delta x \Delta p \sim h$ 。盒子动量不确定度又来源于光子逸出造成的冲量的不确定，而冲量不确定来自光子质量 m 的不确定，从质能关系 $E = mc^2$ 可知， m 的不确定本质来自光子能量的不确定量 ΔE 。另一方面从广义相对论可知，引力势能低处会产生红移和时间变慢。光子高度的不确定 ΔH 就是指针位置不确定值 Δx ，所以 Δx 导致的引力红移（即时间变慢），会造成光子从小孔射出的时间的不确定量 Δt ，而且计算可以证明，此 Δt 与光子能量的不确定量 ΔE ，恰好满足测不准关系 $\Delta t \Delta E \sim h$ 。

爱因斯坦听了玻尔的上述答复后不再讲话。玻尔运用爱因斯坦自己的广义相对论的引力红移理论，成功地反驳了他的反例。图7-16是他们讨论的时候，两人在思考问题。



图7-16 爱因斯坦与玻尔在沉思

争论还在继续

然而，爱因斯坦等人始终没有服气，量子力学的争论一直持续，爱因斯坦后来说了一句很有名的话：“上帝是不掷骰子的”，不可能最后的结果就只是你们的那个概率，概率性的理论不可能是最终理论。他还说：“我花在光量子上的时间是花在广义相对论上的100倍，可还是不知道什么是光量子。”这是爱因斯坦一直到死对这个问题的看法。

著名的物理学家费曼是非常聪明的一个人，他也写过一段话，说：“有人告诉我说他懂得了量子力学，他错了，我相信现在世界上没有一个人真正懂得了量子力学。”

今天玻尔的理论已占了上风，大家都得承认。不管你懂不懂，反正用量子力学概率解释算出的结果跟实验相符。物理学是一门实验和测量的科学，它只承认与实验相符的理论，即使它很难理解。在哥本哈根学

派理论占了统治地位的今天，玻尔强调：“新理论被接受，不是因为反对它的人改变了立场，而是因为反对它的人都死了。”

索尔维会议

图7-17是第一届索尔维会议的照片，这里面都是重要人物。这次索尔维会议是爱因斯坦跟庞加莱见面的唯一一次，坐着的右边第一人是庞加莱。庞加莱是卓越的数学家，同时也在法国的大学里讲理论物理。国外最优秀的教授都是讲课的，不像我们国家，稍微伟大一点儿的人就不想讲课了。坐着的第二人是玛丽·居里，居里夫人，他们是德高望重的学术界老前辈。站立的右边第二人是爱因斯坦。



图7-17 第一届索尔维会议

图中从左到右，坐者：能斯特、布里渊、索尔威、洛伦兹、瓦伯、佩兰、维恩、居里夫人、庞加莱；站者：哥茨米特、普朗克、鲁本斯、索末菲、林德曼、莫里斯·德布罗意、克努

曾、海申诺尔、霍斯特勒、赫森、金斯、卢瑟福、卡麦林—昂纳斯、爱因斯坦、朗之万

对量子物理就简单介绍到这里。

第七讲附录 玻尔对爱因斯坦光子箱实验的答复

1930年，在布鲁塞尔召开的第六届索尔维会议上，爱因斯坦提出著名的“光子箱”思想实验，用来否定“时间能量测不准关系”。

爱因斯坦设想，把箱上小孔的快门打开，其间只让一个光子逸出。由于光子逸出而造成的箱子重量变化，将使箱子上方弹簧秤的指针发生移动。箱子质量的变化等于逸出光子的质量，从质能关系可得到光子的能量。因此，从弹簧秤指针的变化，即可确定逸出光子的能量。而光子逸出的时间即快门打开的时间，可由箱子里的钟准确测定（见图7-15）。

由于光子能量是由弹簧秤测定的，逸出时间是由钟测定的，这两个操作毫无关联，应该能够分别精确测定。这样，光子的能量和逸出时间就同时精确测定了，时间能量测不准关系则不再成立。

爱因斯坦认为，这说明测不准关系有问题，玻尔他们对量子力学的解释不自洽。

玻尔经过一夜的苦苦思索，终于找到了问题的答案，在第二天的会议上，玻尔做了如下答复：

光子的逸出，会使箱子受到一个向上的冲量，这一冲量是由逸出光子的重量 $\left(\frac{E}{c^2}g\right)$ 和逸出过程的时间 t 决定的。此冲量转化为箱子向上的动量 p ，显然

$$p = \left(\frac{E}{c^2}g\right)t \quad (7.8)$$

根据量子力学，箱子的动量和弹簧秤指针显示的箱子的位置 x 应满足测不准关系式（7.6）

$$\Delta x \Delta p \sim h$$

其中

$$\Delta p = t \frac{\Delta E}{c^2} g \quad (7.9)$$

另一方面，根据爱因斯坦广义相对论，时空弯曲得厉害的地方，时间会走得慢，即有引力红移现象。箱子在引力场中高度的变化（弹簧秤指针指示的变化） Δx 造成的时间变慢为 Δt ，可从广义相对论算出

$$\frac{\Delta t}{t} = \frac{g \Delta x}{c^2} \quad (7.10)$$

把式（7.9）与式（7.10）代入式（7.6）马上得出式（7.7）

$$\Delta t \Delta E \sim h$$

这就是时间能量测不准关系。玻尔用爱因斯坦自己的广义相对论，成功地否定了他的“光子箱”反例。爱因斯坦听了玻尔的答复后不再讲话。

实际上，广义相对论的时间变慢式（7.10），从牛顿引力论也可以近似得出。

若光子在引力场中高度升高 Δx ，光子的引力势能将增加，此增加来源于光子动能 ΔE 的减少，所以有

$$-\Delta E = mg \Delta x = \frac{E}{c^2} g \Delta x \quad (7.11)$$

由于 $E = h\nu$ ，所以有

$$-\frac{\Delta \nu}{\nu} = \frac{g}{c^2} \Delta x \quad (7.12)$$

又由于 $\nu \sim \frac{1}{t}$ ，所以有

$$\frac{\Delta t}{t} = \frac{g}{c^2} \Delta x$$

此即引力场中时间变慢的式（7.7）。

第八讲 比一千个太阳还亮



绘画：张京

这一讲我们来介绍核能的利用，特别是原子弹和反应堆。首先讲一下中子的发现。我们已经讲过了元素周期律的发现、光谱线的发现、X射线的发现、天然放射性的发现和量子力学的建立。量子理论发展的同时，核物理的研究也在进展，首先是实验方面。

1. 中子的发现

卢瑟福的猜想

1920年左右，卢瑟福有一个猜测，觉得原子核里，除去质子以外，还应该有一种粒子，质量跟质子差不多一样，但是不带电，也就是我们今天所说的中子。他为什么这样猜测呢？他是根据对元素原子量和原子序的分析，比如说，氦元素的原子量是4，而原子序数是2，也就是说，有两个质子，似乎还有两个与质子质量相似，但是不带电的东西。于是他就产生了可能存在中子的猜测。

卢瑟福的学生当中，有人就开始寻找，但没有找到。1930年，普朗克的研究生玻特，用 α 粒子轰击铍，打出了中子。玻特当时觉得是一种穿透力很强的不带电的射线，以为是 γ 射线。他把这个工作公布了。第二年，约里奥·居里夫妇，对这个问题又进行了研究。

小居里夫妇

约里奥·居里夫妇（图8-1）就是居里夫人的大女儿和大女婿。约里奥出生于一个无产阶级家庭，他祖父是钢铁工人，父亲是巴黎公社社员。巴黎公社失败时他父亲突围，跑到卢森堡，后来平静下来以后，又返回了法国。



图8-1 约里奥·居里夫妇在实验室中

约里奥本来跟居里夫人家没什么关系，原本不大容易接触到居里夫人。但是，他有一个同学，也是他的朋友，是郎之万的儿子。他们两人，一块儿上学，后来又在“一战”中一起当兵。战争结束，两人复原后工作不好找，郎之万的儿子就上他父亲的实验室去了。约里奥找不着工作。郎之万的儿子跟他父亲讲，说你看我这个好朋友找不到工作，咱们

这儿是不是还需要人。他父亲说我们这个实验室已经满了。他又问那你能不能问问居里夫人，看看她那个实验室还缺不缺人。结果他父亲真去问了居里夫人，居里夫人说那让他来吧。来了一谈话，觉得这小伙子还行，就留下来了，于是约里奥就在那儿工作了。

这个约里奥是学化学出身的。居里夫人的大女儿伊琳·居里比他大两三岁吧，也在那个实验室工作。伊琳·居里从小没有上正规学校，而是由几位科学家轮流讲课培养出来的。当时居里夫人等几个科学家搞教改试验，不让他们的孩子上学，由他们几个人亲自出面来讲课，培训这些孩子。伊琳主要学的化学，就在实验室帮她妈做实验。她的妹妹，是学音乐的，后来给她妈写了本传记，就是著名的《居里夫人传》，忠实地记录了她母亲艰苦奋斗的一生。

老居里夫妇就有两个女儿。伊琳这个人很文静，不爱讲话，而约里奥呢，特别爱讲话，于是就产生了一种吸引力。他们晚上做实验总是做到很晚，约里奥经常送伊琳回家。在居里夫人的同意之下，两个人谈了恋爱，并结婚了，这就是著名的约里奥·居里夫妇。两个人后来在实验上，表现出很强的能力。

中子的发现

约里奥夫妇对玻特发现的射线很感兴趣，就用这种射线打击石蜡，从中打出了质子。他们觉得这可能是 γ 射线。他们做了详尽的实验进行研究，但他们头脑中没有中子这个概念。实际上，从动量守恒和能量守恒来看，用 γ 射线不可能打出质量那么大的质子来，但是他们物理略逊一筹，而且头脑中没有可能存在中子的想法，没有这种思想准备。

约里奥夫妇的论文一出来，卢瑟福的学生查德威克看到了，他正在那儿找中子。查德威克高兴极了，哎呀！他们看见了中子还不知道！挺高兴。于是，他也设计了一个跟那个实验类似的实验，把结果登在

《Nature》（《自然》杂志）上面，题目是《中子可能存在》。接着又登了一篇长文章，在英国皇家学会会刊上登出，题目是《中子的存

在》。这下中子就被发现了。

约里奥夫妇很懊丧，自己做出来的发现，就从自己手底下溜走了。这正应了法国著名的生物学家巴斯德的一句话：“机遇只偏爱有准备的头脑。”没有准备的头脑，机遇就会错过去。实际上，人的一生当中都会有很多机遇，大部分、绝大部分都被滑过去了。一旦抓住，就有可能做出大成绩。

约里奥夫妇虽然沮丧，但没有停止科学探索。他们两个人继续研究。不久，就用人工的方法制造出了放射性元素。在此之前放射性元素都是天然的，他们最先用人工的方法制造了放射性元素。

1935年，诺贝尔奖评委会认为中子的发现应该发奖，但是有争论，有人认为，应该由查德威克和约里奥夫妇分享这次的诺贝尔奖，但是这个委员会的主席是查德威克的老师卢瑟福，卢瑟福说：“约里奥夫妇那么聪明，他们以后还会有机会的，这次的奖就给查德威克一个人吧。”当然他也有一定的道理，因为还有玻特呢，一次奖最多只能发给三个人呀。

当年的下半年，同一个评委会评化学奖，因为物理奖和化学奖是同一个评委会评的。大家一致同意把化学奖给约里奥夫妇，理由是他们发现了人工放射性。其实大家也在想，中子的发现他们也是有贡献的。玻特后来因为研究宇宙线获得了诺贝尔物理奖。大概评委会也考虑了他对中子的发现也是有贡献的。大家都得了奖了，应该说最后还是比较公平的。

2. 裂变与链式反应

裂变的发现

现在来讲裂变的发现。1938年，约里奥夫妇用中子轰击铀，发现似乎生成了镭这种元素。在此之前发现的放射性，都是放出一个质子，原子序数减少1，放出一个 α 粒子，原子序数减少2，发射一个电子，原子

序数增加1，反正原子序数只改变1或2，这回一下子从92似乎变成57了，此外还产生了一大堆其他东西，一时也不清楚都是什么。当时，约里奥夫人，即伊琳·居里，在实验室中宣布了这个结果，钱三强亲耳听过她这个报告。但是他们没有弄清楚这是怎么回事，觉得非常奇怪。

这个消息传到德国，德国有一个研究核物理的实验室，其中，有一个化学家叫哈恩，哈恩他们重复了约里奥夫妇的实验，肯定了产物就是镧和钡，他也觉得很奇怪。他们实验室原本有一个女物理学家，但是由于希特勒的迫害已经离开了，因为她是犹太人。这个人就是迈特纳。迈特纳搞物理，哈恩搞化学，两个人关系还可以，合作得也不错，工作做得很出色（图8-2）。

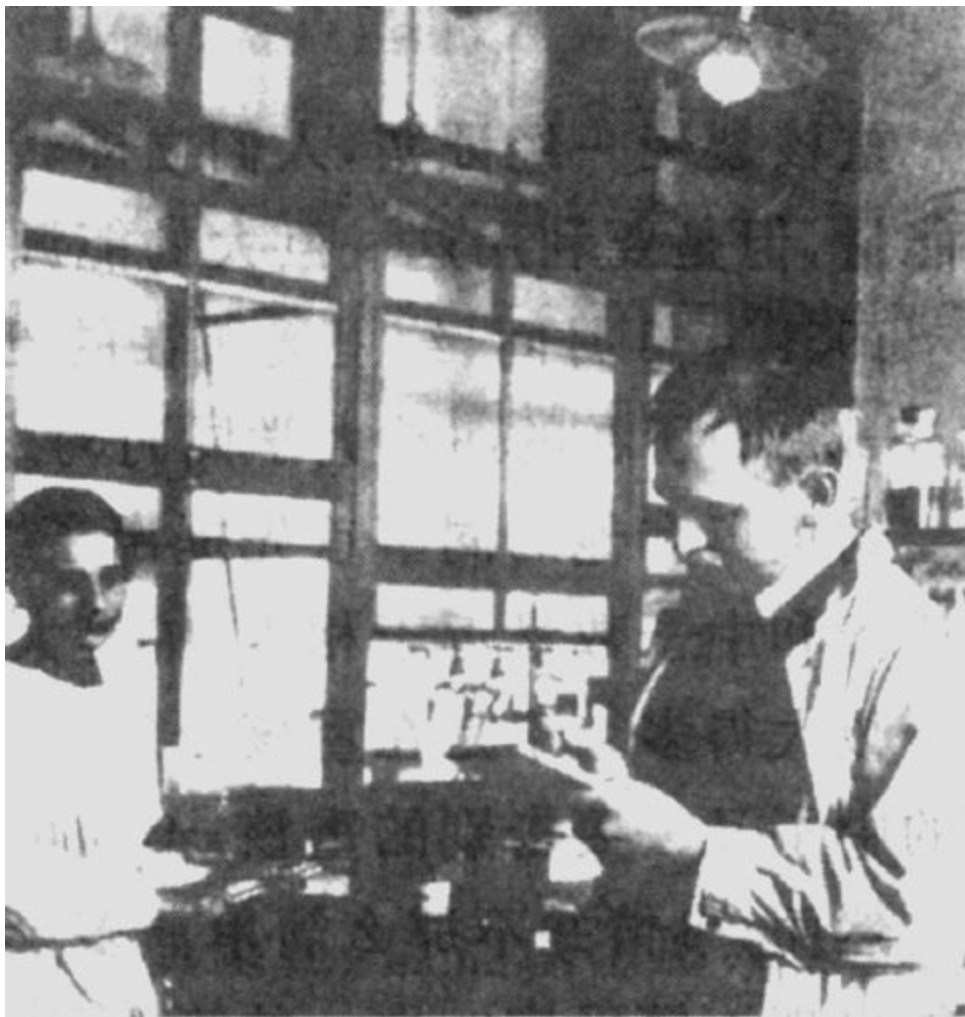


图8-2 哈恩和迈特纳在做实验

迈特纳家的人都非常聪明。她由于是犹太人，感到希特勒上台以后迫害越来越厉害，于是就离开了这个实验室，流亡国外。她有一个侄子，叫弗里希，也是个核物理学家。他走得比较晚，想走的时候已经来不及了，希特勒准备把德国的犹太人都消灭在德国，不允许他们走了。

这时，恰好玻尔到那儿访问，玻尔跟弗里希单独在一起的时候，悄悄问他：你需不需要什么帮助？弗里希说我想赶快离开德国，你能不能帮我。玻尔回去以后，给弗里希发了一封邀请信，邀请他到哥本哈根做短期访问。纳粹官员觉得很为难，玻尔威望那么高，你不让他去也不好。又一想，反正是个短期访问，还得回来，就让他去了。结果他一去就不回来了。

核的液滴模型

1939年的新年，弗里希跟他的姑姑迈特纳两个人，在瑞典共度新年。犹太人不过圣诞节。因为耶稣虽然是犹太人，但他是犹太穷人的领袖。一般的犹太人不承认天主教，他们信犹太教，不承认耶稣，不承认耶稣的神圣地位。

迈特纳他们滑雪回来的时候看到了哈恩的信。看过信以后，对他的实验结果非常惊讶。弗里希觉得这根本就不可能，肯定是哈恩的实验做错了，迈特纳说不会的，我跟哈恩合作多年，他的实验技能非常精细，非常可靠，这个实验一定是可靠的。如果是这样的话，就好像是一个铀核分裂成了大小相近的两块。

于是迈特纳他们就构造了一个液滴模型，铀核像液滴一样，然后有可能变形、拉长、最后分裂（图8-3）。他们用这样的模型，在理论上对核分裂做了解释。顺便提一下，咱们国家的钱三强、何泽慧曾经发现了核的三分裂和四分裂。铀核分裂会有能量放出来，但是仅仅一个单独的铀核裂成两个，能有多大能量，不可能有太多的能量，在工业上无法

利用。

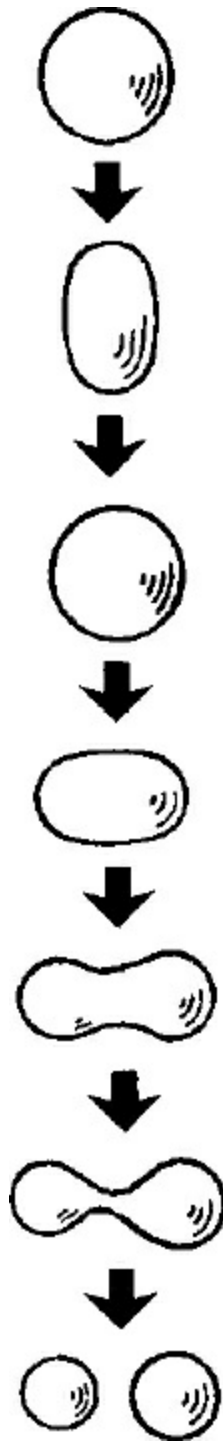


图8-3 核裂变的液滴模型

链式反应

这时候约里奥·居里又做出一个重大的发现，约里奥·居里发现重核里的中子数是远远大于质子数的。比如铀，它的原子序是92，也就是说原子核中含有92个质子。但是它的原子量是二百多，也就是说中子数是远远大于质子数的。大家知道，氢核有两个质子两个中子。一般轻原子核中的中子数和质子数都基本上是相等的。所以如果一个铀核裂成了两个较轻的核，应该有多余的中子出来，而且这些中子出来以后，还有可能刺激别的铀核分裂，因为他们知道中子是可以刺激铀核分裂的。

如果一个铀核自发地裂变放出了多余的中子，这些中子刺激其他的铀核再裂变，再放出更多的中子，使更多的铀核再裂变，……，这就形成了一种链式反应，如图8-4所示，像雪崩一样，能让核能大量地释放出来。这就可以用于工业生产，也可以用来制造武器了。

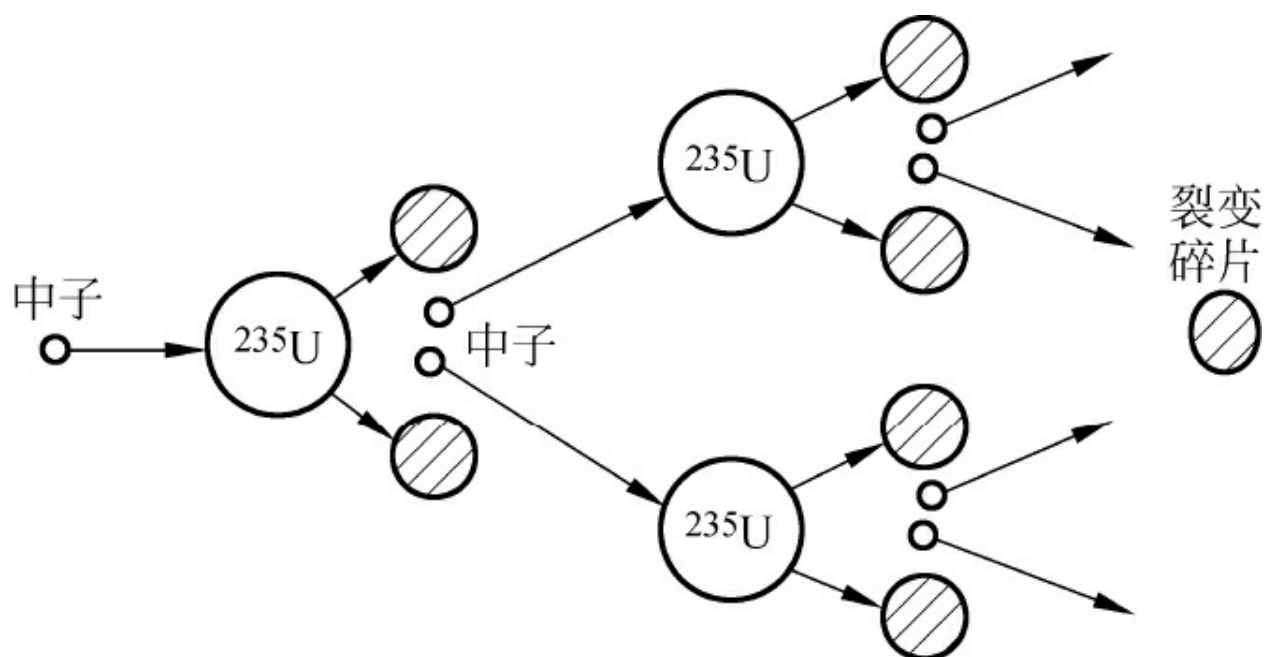


图8-4 链式反应

约里奥很快就用实验证实了自己的想法，他明白自己找到了一条大规模利用核能的途径，找到了一种新的，可以大规模应用的能源。

约里奥立刻把自己的两个助手约到咖啡馆，商量是否公布这一重大发现。当时欧洲上空已经战云密布，他们的发现有可能用于战争，带来

灾难。他们三个人讨论以后认为：火和电的发现都曾经给人类带来过灾难，但更多的是人类文明的进步，应该相信人类可以掌握自己的命运，所以他们公布了自己的发现，把论文登出来了，但是他们没有说出能造原子弹，所以一般人也不大注意。

半年多之后，美国的两个小组，即费米和西纳德领导的两个小组，也分别做出了同样的发现。但是约里奥这个组是最早做出来的。约里奥后来于1944年担任法国科学院院长，1946年担任法国原子能委员会主席。他设计建造了美国之外的第一座核反应堆。他来过中国，到中国来访问，当然还有其他原因，他是法国共产党的中央委员，世界和平理事会的主席。

裂变与聚变

现在解释一下原子核裂变为什么会放出能量。假如有一个原子，它是由 Z 个质子和 N 个中子组成的，那么它总质量是不是就是 Z 乘上 m_p 加上 N 乘上 m_n 呢？这里 m_n 是中子质量， m_p 是质子质量，是不是就是前面这两项相加呢？实验发现不是如此。测得的核质量是式中后面那一项 m ，比前两项加起来要小。

$$B = [Zm_p + Nm_n - m(Z, A)]c^2 \quad (8.1)$$

为什么呢？就是这些质子和中子聚集在一起形成原子核以后会放出一部分能量，放出的这部分能量，就是可以利用的能量。这个差 B 叫做结合能。

许多人研究过单个核子的平均结合能，就是用这种原子核的结合能的数量 B ，除以核中的核子（质子和中子）数，所得结果就是单个核子的平均结合能。图8-5中有一条实验曲线，这条实验曲线显示了各种原子核的平均结合能。

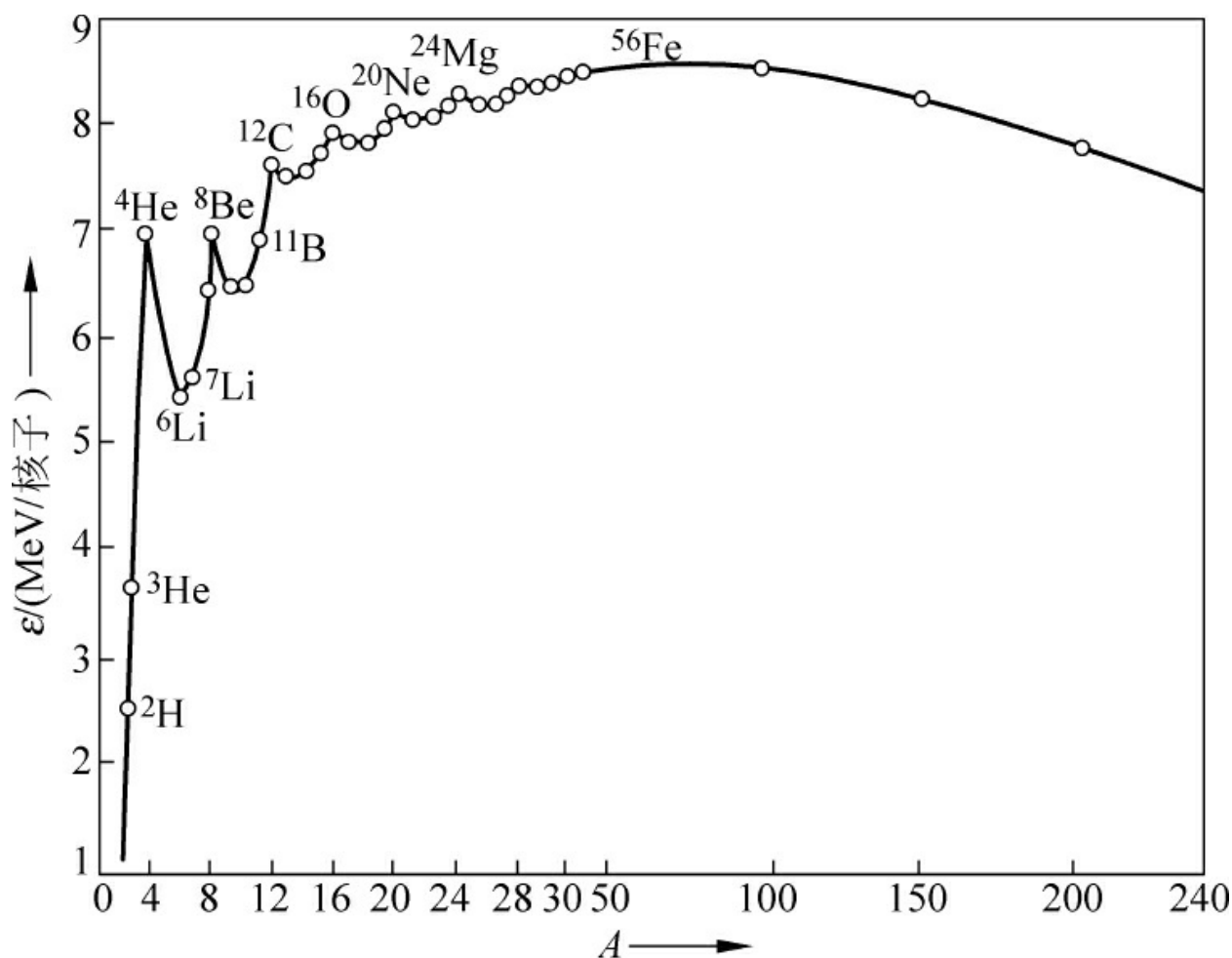


图8-5 各种元素原子核的平均结合能

对于轻元素来讲，它是比较低的，对于特别重的元素来说也是比较低的，但位于中间的元素则比较高。也就是说重的核一旦裂变，分裂为中间这些小的、比较轻的核，就会有能量放出来。因为中间的这些核的平均结合能比重核高。另一方面，中间的这些核的平均结合能比轻核也要高，所以轻元素的核聚合在一起也会放出能量，一种是裂变，一种是聚变，都可以获得能量。

汤川的交换力

日本的汤川秀树是亚洲第一个得诺贝尔奖的人。近代史上，日本人对中国人的鼓舞还是很大的，他们有几个人做出很大的成绩，当然华人就会想，他们和我们同文同种，他们能做，为什么我们不能做，对不

对。

当时猜测原子核内部有一种很强的引力，否则的话，原子核就会分裂了。这种力在原子核内部应该很强，应该远强于质子所带正电荷间的斥力。但是它的作用半径应该比较短，超出原子核之外就没有了。这种力就叫做核力。

汤川秀树推测这是一种交换力。中子和质子之间，或者质子和质子之间，中子和中子之间，交换一种粒子，交换这种粒子，就会产生一种吸引力，产生一种很强的吸引力。他利用测不准关系和质能关系，也就是利用下面这两个关系，估算出了这种交换粒子的质量，是在质子和电子之间，所以叫做介子。实验证实了确实有介子存在。近年来，物理学家把汤川的理论发展改造成夸克理论和量子色动力学。你们回去可以利用这两个关系自己推算一下介子的质量。

$$\Delta t \Delta E \sim h \quad (8.2)$$

$$\Delta E = \Delta m \cdot c^2 \quad (8.3)$$

3. 科学中心向美国迁移

约里奥首提原子弹

接着讲原子能的利用。首先是用在军事上。1939年，德国进攻波兰，欧洲战场全面爆发，法国对德宣战以后，约里奥找了法国军备部部长，建议法国制造原子弹对付希特勒。但是，法国没能挺住，马奇诺防线很快被突破了。

“一战”的时候，德国和法国打阵地战，死的人很多。德国人总结出一套理论，说甭管对方的工事有多厚，有多结实，只要我们的大炮造得够粗，就能把它轰开。法国人发展了一套相反的理论，说甭管对方的炮有多粗，只要我们的工事修得够厚的话，它就轰不开。

所以“一战”结束以后，法国把它的军费都用在修马奇诺防线，一个固定的很牢固的防线上，以为这样就可以在未来的战争中挡住德国人。

结果希特勒的军队绕过了马奇诺防线，法国军队一下子就垮了。这样法国就没法造原子弹了。

约里奥在紧急情况之下，把他的实验室中的重水押运到法国南部的一个港口，装上了一艘去英国的船。然后他就返回巴黎，回到巴黎的时候，巴黎已经被德国占领了。德国占领军传讯了约里奥，问你的重水哪儿去了？约里奥在回巴黎的路上就听说从那个港口开出的另外一条去挪威的船被炸沉了，他就说放在那条去挪威的船上了。那些重水是挪威的，他打算还给挪威，就在那条被炸沉的船上。德国人也就没有再问。因为德国当时已经普及初中教育了，德军官兵都知道居里这个伟大的名字，谁也不想没事给他们找太多的麻烦。德国人没有再找他更多的麻烦，但他也没法造原子弹了，只能搞些一般的科学研究。他后来参加了地下的法国共产党，在实验室中给游击队造炸药。“二战”快结束时，约里奥还拿着手榴弹参加了解放巴黎的最后的巷战，在“二战”期间约里奥夫妇大概就处在这么一种状态。

这个时候，在美国的科学家开始建议美国造原子弹了。他们很担心希特勒能造出来。因为德国有海森堡，有哈恩，搞实验的和搞理论的都有，而且海森堡纳粹思想非常严重，吹捧希特勒非常厉害，大家觉得他们肯定会帮助希特勒造原子弹。希特勒如果造出来，大家都毫不怀疑他肯定敢用，他绝对不会有什么顾虑的。

费米——物理的全才

此时，已有一大批欧洲的物理学家逃到了美国，其中一个费米。费米是意大利的物理学家。他是在相对论、统计物理、核物理这些方面，都做出了重大贡献的物理学家，而且理论和实验都行，后来设计了世界上第一座原子反应堆。他还培养了六个得诺贝尔奖的学生。所以费米是非常了不起的科学家。

在意大利，费米刚开始冒出来的时候，自己并没有太多的信心，他觉得意大利的科学不行。他说，在一个全是聋子的国家里面，有一个人

有一只耳朵好用，大家就觉得这个人听力好得不得了，但是别的国家的人可能都有两只耳朵呀，你要去比就会觉得不行了。后来他有一次到德国去访问了几个月，访问后他觉得自己还是属于有两只耳朵的人，并不比德国人差。

费米后来做出了很多成绩，并获得了诺贝尔物理学奖。他去领奖的时候，准备趁此机会逃离意大利。他本人并不是犹太人，但他的夫人是犹太人，费米夫人跟他从小青梅竹马长大。费米带着全家到瑞典去领诺贝尔奖。那时为了安抚意大利的使馆，不要让他们怀疑自己准备走。费米还特地跑到意大利的使馆去问一问，回国的时候要办什么手续。结果当别人都不在场的时候，意大利使馆那位跟他谈话的工作人员就悄悄跟他说：“费米教授，您全家不是都出来了么，您夫人不是犹太人么，你还回去干嘛呀？”于是他就放心了。其实，离开意大利之前，他就去过美国大使馆，要求移居美国。美国大使馆的人不认识他，说我们美国，欢迎欧洲移民。但是那个工作人员又跟费米讲，不过有一条，弱智的人我们不要，所以明天您和您的全家必须来做一下智力检测。做过智力检测后，美国方面突然发现这位是诺贝尔奖获得者，赶紧就说不用再等什么了，马上就给你们签证，你们去美国吧。

顺便提一下，当时美国欢迎欧洲移民，不欢迎亚洲移民，美国、加拿大都对不起我们华人，帮他们修了那么长的铁路，最后他们把这些华人都赶走了。你要在这里呆着，可以。但女人进不来，华人的妇女都不能进来，让你们自生自灭。当地人也不嫁给你，连印第安人都看不起华人。最近美国和加拿大政府都曾经为此有过道歉的表示。

费米的夫人回忆过跟费米的交往。他们俩的父亲是好朋友，所以他们从小就认识。费米的夫人说：“我小时候第一次参加舞会时，只有十五岁，我打扮了很长时间，觉得非常高兴，不过当时我还是一只丑小鸭，我一直在那里坐着就没有一位男士来请我跳舞，真是非常尴尬。后来终于有个小伙子走到跟前，邀请我跳舞，这个人就是费米。”她说她

非常感激，让她从这种尴尬的局面下解脱出来。在诺贝尔奖颁奖的时候，有一个仪式，就是获奖者的夫人会有机会和瑞典的王子一块儿跳舞。她（费米夫人）说：跟一个白马王子在一块儿跳舞，这是每一个小姑娘从小就有的梦想，这次我终于有了这个机会，不过呢，这位王子已经六十岁了。

炸弹仓里的玻尔

玻尔也移居到了美国。玻尔有犹太血统。“二战”时丹麦对德国是有条件投降。因为开战以后德国的军舰一下就把丹麦唯一的那艘军舰给打沉了。丹麦海军马上就不行了，根本就没多少装备，但是哥本哈根卫戍司令不想屈服，准备把武器发给老百姓进行巷战。国王说算了，咱们根本打不过德国，和德国签订停战协定吧。投降，但是要有条件，其中一个条件就是犹太人你们不能带走。在这一点上，丹麦人还是够意思的。丹麦方面说，你们强迫把犹太人聚集到一起可以，但是要允许他们的邻居去探望他们，因为他们的邻居对他们的安全很关心。所以丹麦的犹太人差不多都活下来了。

当时希特勒在德国屠杀犹太人已经很厉害了，所以玻尔心里很担忧。这时候海森堡跑到哥本哈根去了。在玻尔的研究所附近，一些拥护纳粹的丹麦人建立了另外一个由纯种雅利安人组成的物理研究所，就是没有犹太人的一个研究所。

海森堡到那里作报告，作报告时他还故意问了一句，说玻尔教授怎么没来，还提醒大伙儿玻尔没有来，玻尔不是纯种人。接着海森堡又去拜访玻尔，两个人在实验室里谈话，又在草地上谈话，谈了些什么，后来两个人说的完全不一样。“二战”结束以后把海森堡抓起来了，海森堡说我那时一直劝告玻尔，说你是安全的，绝对没有任何问题等。玻尔说根本不是这么一回事，他认为海森堡是坚定的希特勒的拥护者。

玻尔觉得自己非常不安全，他最后选择了偷渡，坐船逃到挪威，坐一艘小船冒险逃往挪威，海水把他全身都打湿了。逃到挪威以后，他又

坐上一架英国的轰炸机，飞往了英国，然后坐船去了美国。坐轰炸机的时候，轰炸机上每个人都有位置，机枪手你也不能让他腾出位置，为什么呢，因为中途可能碰上德国飞机，机枪手还有任务。玻尔没地方呆，就把他搁到炸弹仓里了，结果玻尔还休克了，因为有一段高空缺氧，他没有戴好氧气面罩，休克了，当然好赖还是到了伦敦。

“二战”以后有些记者，西方的记者就质问英国政府说，听说你们给驾驶员有个命令：这飞机如果要迫降的话，不能把玻尔交给德国人，要把他扔到海里。是不是有这么回事？一直质问，一直到70年代、80年代还有人在那儿质问英国当局，英国当局断然否认有这种事情。后来玻尔终于逃到了美国。

4. 原子弹的研制

美国启动曼哈顿工程

这些逃到美国的科学家都觉得希特勒可能制造原子弹，所以他们就希望美国赶紧造，一定要抢在希特勒之前造出来。于是，几个物理学家，西拉德、泰勒、维格纳，打算写信给美国总统，劝美国赶紧造原子弹。他们觉得自己的影响力不够，就去拜访爱因斯坦，请爱因斯坦写封信，爱因斯坦同意了，在信上签了字。这封信有人说是西拉德他们写的，但是甭管怎么说吧，反正这封信是爱因斯坦签的字，签了字就是爱因斯坦的信。当时有一个罗斯福的朋友萨克斯，拿着这封信去见罗斯福，想跟罗斯福谈谈，但罗斯福没大听懂，罗斯福那时忙得不得了。

当时美国正在备战，因为觉得德国可能马上就会进攻美国，所以罗斯福当时紧张得不得了。美国总统是三军统帅，各种外交活动也非常多。美国当时已经在把大量军火卖给英国，还派舰队给运输船护航，以防德国潜艇攻击。同时允许他们的空军人员退役以后参加陈纳德的志愿空军到中国作战，不过现役人员必须选择退伍才能去。美国和德国、日本的矛盾越来越尖锐。

这时候，科学家们谈造原子弹的事，罗斯福根本听不进去，说我忙得这么厉害，你这个事儿到底有什么用也搞不清楚，就很冷淡。后来罗斯福看这个朋友不太高兴，就说这样吧，明天早上，我请你吃早饭，你还可以再谈谈你那个东西。萨克斯一晚上没睡好。第二天吃早饭，见了罗斯福他就问，你知道拿破仑为什么会失败么？这个问题问得也够远的。罗斯福说为什么会失败，他说就是因为拿破仑当时没有相信先进的科学技术。

萨克斯说，当时有两个美国工程师向拿破仑建议过，把蒸汽机装到船上，造机器船，这样在逆风的时候都可以在英国登陆。拿破仑就笑了，说他不相信没有帆这船还能走。于是没采纳这两个美国工程师的意见，结果法国就是因为海军干不过英国，最后失败了。

罗斯福说，那照你的意思我要是不造这个大炸弹的话，我也会跟拿破仑一样最后失败了？萨克斯说是这样。好，罗斯福就按了一下电铃，把他的副官，一个将军找来，对他说，现在有一个重要的任务，这位先生会跟你讲，这件事情要马上行动。

这件美国政府称为“曼哈顿工程”的事情刚布置下去，第二天珍珠港事件爆发，日本海军航空兵袭击了珍珠港，把美国炸得够戗。为什么呢，美国当时没想到日本敢这么干，当然也觉得将来日本要打它，但是觉得那还是以后的事情，因此毫无防范。当时美国最担心的主要是德国，所以美国把大的战舰，都从大西洋挪到太平洋来了，移到珍珠港，以为这边安全一点，离德国远点，没想到日本人先动手了，在珍珠港把美国的大战舰炸了个正着，幸亏航空母舰当时不在港里，所以最后还手的还快一点，要不然更不可想象。

奥本海默临危受命

太平洋战争爆发了，美国马上和日本、德国处于战争状态。后来，罗斯福任命熟悉工程建设的格罗夫斯将军负责曼哈顿工程，对他说这件事情你只对我一个人负责，缺钱直接找我，不要跟国会、跟政府的任何

人员谈起，只对我负责。格罗夫斯将军接受了这个使命，开始物色人选。有的人主张找一些德高望重的老教授来做，但是老教授一般都比较保守，而且精力也有限。将军发现有一个40岁左右的中年人比较合适。此人叫奥本海默，是在德国哥丁根大学毕业的。

“二战”前，美国很多人跑到英国和德国去留学，因为美国那个时候科学还不太发达，最先进的还是欧洲。奥本海默就跑到德国去，在哥丁根大学上学。奥本海默在德国表现得很出色。他是玻恩的学生，他经常打断别人的报告，别人在上面作报告，正讲到半截，他就上去了，把人家粉笔拿过来，说这个问题啊，其实根本用不着像你这么讲，你要是这样讲它就简单多了，奥本海默净弄这种事儿。有一次玻恩写了一篇论文，给他看，说过两天你来找我，讨论讨论。过两天他来了，玻恩问他说，这篇文章你看了吗？觉得怎么样？他说这篇文章写得很好啊，真是你写的吗？怀疑他的老师是不是能写出这样的文章来，大概是因为玻恩的论文平时出错的可能性比较高，奥本海默有点怀疑。总之，奥本海默因为这种性格，得罪了不少人。

有些人向这位将军提议，奥本海默这个人够聪明的，可以胜任原子弹的研究。这时候他已经预言了黑洞了，已经预言过暗星了，当时叫暗星，不叫黑洞。但联邦调查局说奥本海默不行，将军问为什么不行，联邦调查局说奥本海默这个人倾向共产党，他弟弟和弟媳都是美共党员，他的女朋友也是美共党员，他老看共产党的宣传品，说他这个人靠不住。可是将军觉得又没有其他合适的人。

比如说找到一个教授，这个人说可以造，一算出来，造出来那么大家伙，跟个小山似的。原子弹最后是要给人家送去的，你造个小山似的，就地引爆它能有多大作用？最后还是奥本海默合适，格罗夫斯将军就跟联邦调查局的人说，你们先别着急，你们把材料拿过来我看一看。他一看，也没有什么。因为职业军人和一般科学家在政治上并不那么敏感，搞情报工作的人把一个人说得黑不溜秋，其实别人看来，也没

有什么太多的证据说明他有什么问题。于是将军就说还是用奥本海默吧。他跟联邦调查局的人讲，你们不用管了，这件事情我直接对总统负责。

奥本海默受命担任了原子弹的总设计师。奥本海默开始找一些人，到他那儿工作。其中很重要的一个任务，是测铀的临界质量。大家知道，原子弹爆炸，原则上来讲，只要这个铀的块头，超过了一个临界质量，它一定会炸。原因是这样的，铀里面一定会（有一定的概率）有一些铀核自发裂变，放出中子。放出的中子，会刺激别的铀核发生裂变，但是如果这个铀块儿不够大，那么很多中子就飞出去了，碰不到别的铀核，也就没有贡献了。再有呢，铀里有杂质，这些杂质会吸收掉一些中子，使它们不再做出贡献。而且因为你造原子弹的铀是工业品，它不可能是绝对纯的，一定有杂质，而且每一次造出来的铀，杂质的成分和含量都会有差别，所以，就要经常的测定铀的临界质量，以便确定多大个的铀块会爆炸。当时要找一个人来测临界质量，这件工作需要一个实验非常精细，又非常勇敢的人来做，需要有牺牲精神的人。

勇于献身的斯洛廷

有人推荐加拿大的年轻物理学家斯洛廷，说这个人可以胜任。他实验做得不错，而且很勇敢，战争一爆发他就要求参军，后来发现他是近视眼，军队不要他。但是测临界质量呢，近视眼没有关系，可以做。于是把斯洛廷找来，他同意了。他把两块铀装在一个架子上，然后用螺丝刀拧，使两块铀靠近，周围放了很多计数器，如果计数器嘎嘎响得厉害了，就赶紧把两块铀再拧开。拧开以后拿刀片削下一块，再让它们靠近，一直到计数器刚刚不响的时候，这时铀的质量就是临界质量，超过这个值它就会自动爆炸了。这样我们就知道，多大块的铀，就可以制成原子弹了。斯洛廷知道自己的工作很危险，他说我现在是在玩龙尾巴。龙在西方是一种很可恶、很凶狠的东西。中国人老说自己是龙，人家听着就有点害怕。

“二战”之后，斯洛廷他们有一次测定的时候出事儿了。两块钚（钚是另一种可制成原子弹的元素）靠近的时候，计数器开始响了，正在此时他的螺丝刀一下掉到了地上，没法拧了，顿时那间屋子全被钚发的光照亮了。他只好马上用手把这两块钚掰开。当时他们几个人是坐成一圈的，斯洛廷说我是活不了多久了，但是你们还可以多活一段时间，然后他就把每一个人的位置，身体的什么部位对着钚块，都画在了黑板上。

5. 爱开玩笑的费曼

保险柜怎么开了？

核基地里有很多很优秀的人，比如费曼，他很年轻就到这个地方工作了。费曼这个人非常聪明，又非常爱开玩笑。他到原子弹试验场后很不习惯，因为美国人都自由散漫惯了，这里的工作要保密，弄得又是进门卡啦，又是什么别的措施啦，每天工作完了的那些资料都要放进保密箱啦，他都觉得很习惯。特别是夫人来信了，还有人要拆开看一下，他很不高兴。

有一次他回去探亲的时候，就一连写了好几封信，把每一封信都撕成碎末，塞到信封里面，然后跟他夫人说，你每过一个星期给我寄一封。信寄到后，这帮联邦调查局的，一查，唉，怎么都是碎末啊。就拼凑，拼了半天也没看出有什么问题。然后又来了一封，又是碎末。

费曼觉得那些人净搞些形式主义的东西，你看，把大门看得那么严，查得那么严，墙上有个洞他们不管。有一次，他就从大门出去，然后从那个洞钻进来，然后又出去，再钻进来，想吸引警卫的注意。警卫刚开始没注意他，后来终于有人注意到了，说这个人怎么只看见他出去，没看见他进来啊，就把他拦住了，这才发现墙上有个洞。

还有一次，他把保密员锁保险柜的密码，给猜出来了。趁着保密员出去，他就把保险柜一个一个都打开，每个保险柜里都搁了一张纸条，搁进去以后，他就躲起来了。那保密员回来一看，哟，这保险柜怎么打

开了，马上就把电铃按响了。保安人员哇地跑过来，研究所里面所有的人都跑出来看，他也跑出来混在里面看。怎么回事啊，一看，保险箱里有个纸条，哦，说看几号保险箱，然后大家就去看那个保险箱，又有一个纸条说看几号保险箱，一直查看，看到最后一个保险箱，箱里的纸条写着：猜猜看是谁干的。

被当做学生的教授

“二战”之后，费曼离开了核实验基地，到康奈尔大学工作。“二战”期间美国动员了一两千万人参军。战争一结束，这么多年轻人往哪儿安置啊！工作安排不了，于是进大学。所以大学里简直人满为患，而且多大岁数的人都有。费曼到那儿去报到。报到的时候正好已经下班了，他想先找一个住的地方，就找到学校宿舍的管理员。那个宿舍管理员说，告诉你啊，年轻人，真的是没有地方。管理员把他当成学生了，他说实在是没有地方，我要是剩有一个地方都让你住，但实在是没有地方了。

费曼没办法，就想在树林里坐一宿，第二天早晨再去找系办公室解决。但是晚上坐得很冷，不得不返回宿舍楼。没办法，只好在宿舍楼走廊里的长椅子上睡了一宿。等到天亮一上班，他就去找物理系主任。系主任说，唉，怎么会没有你的房子，单独给你留了房子啊，他们一定是搞错了，以为你是学生。于是系主任马上打了个电话，然后让费曼赶紧去，说有你的房子。他就去了。去了一看，原来值夜班的管理员下班了，换了个值白班的。值白班的人也不认识他，就对他说：年轻人，我告诉你，真的是没有房子，你知道不知道，昨天有个教授就在那个长椅子上睡了一夜。费曼说，我就是那个教授，我不想在那儿再睡一夜了。管理员恍然大悟，马上给他打开了一个房间。

费曼刚住到房间里，就有人敲门，一开门，是两个高年级的女同学来找他。对他说，你这个岁数上大学呢，是晚了点儿，不过没有关系，如果你有困难的话，我们可以帮助你。费曼说我不是学生，我是教授。

那两个人一听，觉得这家伙是个骗子吧，就走了。因为美国人很讨厌说假话的人。还有一次他跟着大家去参加晚会，有个女生和他跳舞，问他是哪个年级的，他说我不是学生，是教授。那个女生怀疑他欺骗，就说，那你还造过原子弹吧。他说是啊，我是造过原子弹。那个女生把手一甩，说“该死的骗子！”，转身就走了。

“神奇”的制图板

美国的教授还真的什么课都讲，费曼连制图课也上过。他说有些学生学得特别呆板，一点儿不会灵活运用。在制图课上，他拿起制图板，说你们看，制图板上有很多曲线，各式各样的曲线。他说，制图板上的曲线都有一个特点，你不管怎么转，它最低的那一点的切线一定是水平的。于是，他看见所有的学生，几十个学生都把自己的制图板拿起来在那儿转，以为这真是制图板上的曲线的特点。大家知道是怎么回事吗？你们学过微积分，不管这曲线怎么转，这极小值处的导数肯定是零啊，最下面的那一点（极小值）的切线当然是水平的，是不是。他说你看，这些学生学得真是太呆板了。

谁经历的时间最长？

费曼不仅跟学生开玩笑，还跟爱因斯坦的助手开玩笑。他问人家一个问题，是与双生子佯谬有关的一个问题。他说，有几个人，从地面飞到高空，然后落下来。他可以坐火箭上去，然后落下来；他也可以被扔上去然后落下来；或者通过其他各种各样的途径（方式）升上去然后落下来。如果地面上的人看到他们同时出发，又同时落地，费曼问，根据相对论，他们当中谁经历的时间最长。爱因斯坦的那个助手没答出来。其实按照相对论，当然是沿测地线运动的那个人经历的时间最长，就是抛上去，再作自由落体运动，即一直作惯性运动的那个人，经历的时间最长。可是爱因斯坦的这位助手没有答出来。第二天那助手找到费曼说，你怎么跟我开这个玩笑啊，那人回去思考了半天，也许还和别人讨论过，最后才明白过来。费曼的意思就是说有些人学得太死，爱因斯坦

的这位助手自己就是搞相对论的，这个问题他居然答不出来。

费曼图和路径积分

费曼这人最大的贡献是发明费曼图。基本粒子之间的相互作用计算起来非常困难，他发明了一种图，大大简化了复杂的计算。费曼太聪明了，他创造那种图（费曼图）的时候，奥本海默那么聪明的人都没看懂。

奥本海默在“二战”结束以后，离开了核基地，出来主持非军事的理论物理研究。“二战”结束以后奥本海默曾经光荣了一段时间，因为他是原子弹之父。但是很快就不行了。因为美国的原子弹的机密泄露了。联邦调查局就说，你看看，我们早就说奥本海默靠不住，肯定是他把秘密泄露给苏联了。苏联甚至还搞到了一块原子弹的样品，肯定是奥本海默干的。于是就向所有的人找证据，要把奥本海默从原子弹试验场弄出去。但又找不到任何证据。在持续的政治迫害下，奥本海默最后终于不得不离开了原子弹研究基地。费曼等很多人都跟着奥本海默一起离开了，到奥本海默主持的那个研究机构去工作。多年后证明，奥本海默是清白的，原子弹机密不是他泄露的。

费曼发明了费曼图，但是没有人能看懂它。直到有一次，另外一位同样年轻的物理学家，和他一起去开会，走到半路上被洪水给挡住了。两个人只好在一个县城的旅馆里住下，没有地方可去。那个人就说你再给我讲讲你那个图吧。他又讲解了一番，那个人终于听懂了。回去以后就对奥本海默说，费曼讲的是对的，费曼图是对的。

费曼还有一个重大创新——路径积分。根据量子论，基本粒子是没有轨道的，电子从一点运动到另一点，没有轨道。但是狄拉克有个思想，认为没有轨道等价于有无穷多条轨道，就是说从一点到另一点，所有的轨道都有贡献，包括超光速的轨道也有贡献。把这些贡献都加起来，最后的结果，跟没有轨道是一样的。狄拉克曾经提到过这个思想，但是他没有往下做。费曼往下做了，而且做成功了，这叫路径积分。

狄拉克比费曼岁数大很多，有一次，他们俩人见面了，这是“二战”结束以后很久的事。费曼见到狄拉克非常高兴，就不断地跟狄拉克讲他工作的成果，哎呀你看我干的这个东西都是在你工作的基础上做的，怎么怎么样跟他讲。狄拉克就靠在那儿一言不发，图8-6是当时有人拍的照片。费曼讲了大概一两个小时，狄拉克一言不发，最后狄拉克终于发言了，他说你等一下，我有一个问题。费曼一听，有个问题，高兴了，要互动一下了。马上就问狄拉克，什么问题。他说：洗手间在什么地方？



图8-6 费曼与狄拉克在交谈

6. 广岛与长崎的蘑菇云

枪法与内爆法（图8-7）

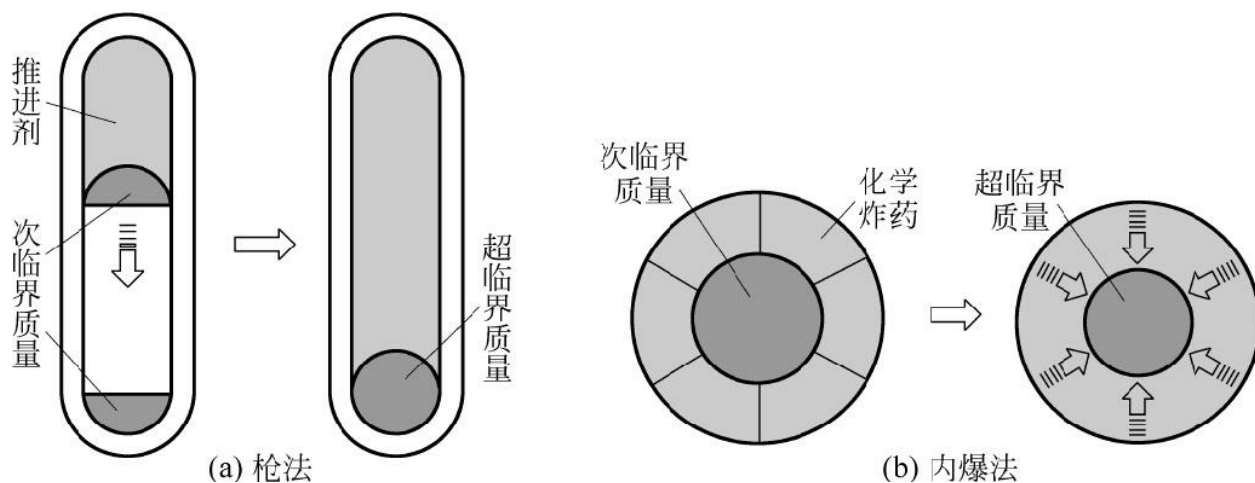


图8-7 枪法与内爆法

我们再来看原子弹。原子弹引爆有两种办法，一种引爆法是枪法。就是两块铀，上面放一点推进剂，就是炸药，这两块铀的任何一块的质量都小于临界质量，炸药一炸，使两块铀合到一起就超过临界质量了，原子弹就爆炸了。还有一种呢，是铀在中间，周围一个球面上，全是炸药，炸药一炸铀块就往中心挤，一挤紧，铀之间的空隙就小了，那么临界质量的要求就低一点，然后就引爆了。这种方法叫做内爆法。

美国1945年7月16日，试爆了第一颗原子弹，在新墨西哥州，这是一颗采用内爆法的钚弹。当时德国已经投降了，美国准备对日本使用，因为美军跟日本在海岛争夺战中死的人太多了，包括硫磺岛等很多岛。美国人在有些岛准备拍电影，拍美国军队怎么迅速占领这个岛，结果这个岛，打了一两个月也打不下来，日本军队几乎没有人投降，都打到底。美国人觉得伤亡太惨重了。

广岛

1945年，美国人决定要使用原子弹来迅速结束战争。讨论是否使用

原子弹的时候，奥本海默不同意用。奥本海默认为现在战争已经到最后了，原子弹分不清平民和军队，这样是不人道的，不同意用。但是美国政府，像那个将军，整个战争期间什么都没有干，就主持造了个原子弹，现在要不用一下也是很遗憾的事儿。所以美国人就准备用原子弹。他们选择了四个日本的城市，基本没有被同盟国军队轰炸过的城市，要试一试原子弹的威力。其实，广岛是其中的第二个目标，并不是首选的那个城市。

1945年8月6日，美国的一架轰炸机带着原子弹去了，正好第一个目标那里是阴天，看不见下面。于是就奔第二个目标，这个目标就是广岛。广岛本来也阴天，正好一阵风把云彩吹开了。飞机到那以后就把原子弹抛下去了。日本是拉了防空警报的。后来看见就一架美国飞机，而且转了一下就飞走了。只留下一个降落伞，吊着个东西。日本人以为那是个气象仪器，认为美国飞机是气象飞机，就把警报解除了。很多日本人就从防空洞中出来，这时候原子弹爆炸了（图8-8），这是一颗用枪法引爆的铀弹。我看过一个日本记者描写的当时广岛被炸的惨状，确实很厉害。



图8-8 广岛的核爆炸

长崎

1945年8月9日，又在长崎扔了一颗，是用内爆法引爆的钚弹。这个原子弹扔的时候苏联已经对日宣战了。苏联原来答应各同盟国的条件就是，德国投降之后不超过三个月，苏联对日宣战。5月9日，德国正式投降。所以8月8日下午，苏联通知日本，从8月9日0点开始，苏联与日本处于战争状态。8月9日呢，美国又在长崎扔了一颗原子弹。现在有很多

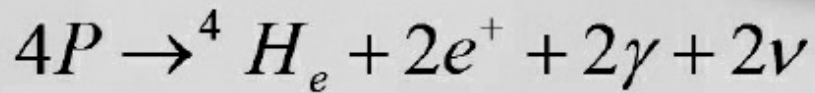
争议说该不该扔这颗原子弹，当然各人都有自己的理由。一方面确实日本死了很多人，但是呢，战争如果不尽快结束那不是中国人也在死么，包括美国人都在死。日本的普通炸弹、炮弹、子弹不也在杀人吗？日本军队不是一直在屠杀平民吗？所以对这件事情会有不同的看法。

泰勒与氢弹

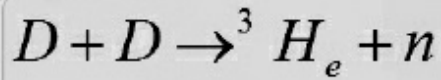
原子弹造好后，又开始造氢弹。氢弹是由泰勒设计的。泰勒是杨振宁的博士生导师，杨振宁说，泰勒这个人有个特点，他一天能出十个主意，其中九个半都是错的。但是那半个对的，就会对工作有点促进。泰勒本来在奥本海默手下工作，他所在那个组的组长来找奥本海默，说你赶紧把这个人调走，这个人一会儿一个主意，昨天我们刚商量好一个方案，他也同意了，今天刚动手做，他又说不行了，这样下去没法干，你干脆把他调走。于是奥本海默就把泰勒叫来，对他说，现在有一件重要工作，需要有一个能干的人单独去完成，我看你比较合适。这件工作就是研制氢弹。泰勒就去了，当时泰勒不知道是领导嫌他碍事。后来泰勒明白了，这事儿实际上是有人对他有意见。

图8-9给出的就是氢弹的原理。当然真正的氢弹的详细原理是比较麻烦的，现在我们给出的是一个示意图。就是质子，结合成氦4。但是实际上，往往不是用通常的氢，而是用氘（氢的同位素，由一个质子和一个中子组成），或者氚（氢的另一种同位素，由一个质子和两个中子组成）来进行。

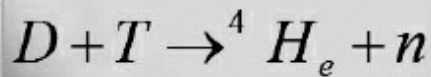
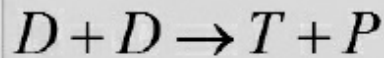
太阳能：



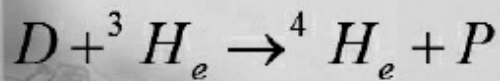
氢弹：



D：氘



T：氚



上亿度高温和高压

图8-9 聚变原理

7. 中国研制核武器

为了自身的安全

下面讲一下中国的原子弹研制。其实在“二战”结束以后蒋介石就想造原子弹，他派了一些人到国外去考察。但是国民党政权很快垮了。中国共产党刚开始掌握政权的时候并没有想造原子弹，这个想法是朝鲜战争导致的。

朝鲜战争进行期间，美国的报纸上就公开讨论对中国使用原子弹的问题。这件事情谁也不敢说它到底是不是准备扔。而且英国首相听美国扬言要扔原子弹以后立刻飞到美国，告诉美国总统，千万别用啊，说你要一用，苏联就会用，当然苏联原子弹扔不到美国，但是还是扔得到我们英国的。中国领导人肯定也知道这种情况，但是苏联安慰中国说，没关系，我们有原子弹。但中国人觉得，你们有总不如我们自个儿有使起来顺手，还是自己拥有比较可靠。

钱三强、杨承宗与约里奥·居里

当时中国从国外回来一些核物理学家，如钱三强、何泽慧，从法国回来了，从居里实验室回来了。钱三强回国是在1946年左右，日本投降以后他就回来了。他们夫妇临离开法国的时候，约里奥曾经对钱三强讲，说你们两个人都是优秀的青年物理学家，我本来希望你们终生在居里实验室工作，但是我也知道你们很热爱你们的祖国，我也不能强留你们。随后，在告别宴会上约里奥拿出一个盒子，说我没有礼物可以送给你们，这里面有些标定了放射性强度的放射性元素，这对你们的国家可能是有用的，就送给钱三强了。钱三强回来以后把它搁在中国科学院的仓库里，后来中国开始核物理研究的时候，把这些东西拿出来用了。钱三强先生的一篇回忆录里讲到这件事情，这对中国还是有一些用处的。

再比如说杨承宗也回来了，他是搞放射化学的。朝鲜战争爆发以后，他想回来，钱三强跟他说你现在别回来，现在国内很乱，根本没有办法开展工作，你等一等再回来。他是约里奥夫人的学生。最后，钱三强还托杨承宗，把中国的一些经费交给了约里奥，请他帮助中国购买核仪器，发展一般的科学研究，当时还不是想制造原子弹，只是想买一些仪器，开展核物理研究。因为美国对中国禁运，什么都不让卖给中国。后来杨承宗要回国了，约里奥夫人说你再去跟我的丈夫告别一下。约里奥当时在另一个搞核物理的研究所工作。约里奥夫人的研究所是搞放射化学的。杨承宗去见了约里奥，约里奥就对他说，请你转告毛泽东，中国要想制止核战争的话，就应该自己拥有核武器。其实原子弹的理论也不是美国人首先创造的。你们中国不是也有自己的科学家么。第一个说出原子弹原理的人其实是约里奥。他不见得告诉咱们什么秘密，但是他说了这番鼓励中国人的话。

回国后，杨承宗把约里奥的话转达给钱三强，钱三强又转达给周总理，一直转达到党中央。我想这件事情大概也使中国的领导人觉得有了

信心，像约里奥这样的专家都认为我们能造原子弹，我们应该相信自己的能力。

中国的“奥本海默”——邓稼先

邓稼先是在那之前回来的。周光召和黄祖洽则是中国自己培养的。邓稼先回国以后在北京大学工作。后来动员他搞原子弹研究的时候，组织上找他谈话，征求他意见。那天晚上他怎么都睡不着，他的夫人问他说，你怎么了，他坐起来说，现在组织上找我谈话了，要我去干一件非常重要的工作。如果我去干呢，咱们俩就会长期两地分居，你不会知道我在哪儿，也不会知道我在干什么，但是可以通信。这件工作如果做成，对咱们国家是非常重要的，对我也是非常重要的。他夫人就说，既然对咱们国家，对你本人都那么重要，那你就去干吧。当时，中国知识分子都非常爱国。他的家人真的一直都不知道他在干什么。中国第一颗原子弹爆炸的时候，他岳父还问郭沫若，说，哎呀，这原子弹造得太棒了，也不知道谁是设计这个原子弹的人。郭沫若说，你去问问你女婿是谁设计的。

美国一直怀疑中国造原子弹是有苏联帮忙的，另外还可能有其他的人，比如美国有一个学核物理的女研究生叫寒春，在中国帮助搞农业机械化，这个人是费米的研究生，参加过美国的核试验。美国人猜测是不是他们帮助了中国。

邓稼先是杨振宁从小的朋友，是杨振宁的同学。杨振宁回国的时候提出来要见邓稼先，周总理就把邓稼先从核试验场叫回来。杨振宁谈话当中总是问，“也不知道谁是中国的‘奥本海默’”，“不知道中国造原子弹有没有外国人帮忙”，邓稼先都把话给岔开了。后来杨振宁路过上海准备回美国，邓稼先向周总理汇报了跟杨振宁的谈话。周总理说，你可以告诉他，没有外国人参加。于是他就写了一封信给杨振宁，总理派了一个信使，坐飞机把信带到上海。杨振宁正在告别宴会上，他就把那封信交给了杨振宁。信里面说，中国只在原子弹研制的初期得到过苏联的一

些帮助，后来就没有了。也没有其他外国人帮忙。当时杨振宁非常激动，立刻就到洗手间去了，他不想让人看到他的眼泪。

中国的核武器的研制，主要依靠自己的力量。寒春并没有参与。因为寒春在参加美国原子弹的研制以后，就觉得不应该再制造杀人武器，她后来搞农业机械化了。苏联后来在技术上对中国是完全封锁了，但是刚开始咱们确实是从苏联得到了一些资料，也得到了苏联的一些帮助。

当原子弹还在研制时期，中国的氢弹研制就开始了。氢弹的设计师是于敏，他是北大毕业的。图8-10是中国的第一颗原子弹爆炸的景象。



图8-10 中国的第一次核试验

8. 和平利用核能

费米创建的第一座核反应堆

第一座核反应堆，是费米在美国芝加哥大学研制出来的，在“二战”期间就已经开始运作了。大家请看图8-11，这张图是后来的人画的，当时不准拍照。大部分人都在核反应堆的左边看，右边还有三个

人，这三个人是“敢死队”。为什么呢，如果这个反应堆要控制不住了，他们抱着一些能够吸收中子的溶液，可以倒进管道里去，让反应停下来，但是如果核反应停不住，这三个人就完了，就肯定不行了，来不及跑了。实验成功之后，他们就打电话向美国当局报告，说“那个航海家已经登上了新大陆”。“当地居民呢”，那边问，他们说“十分友好”。这是暗语，意思是实验成功了。

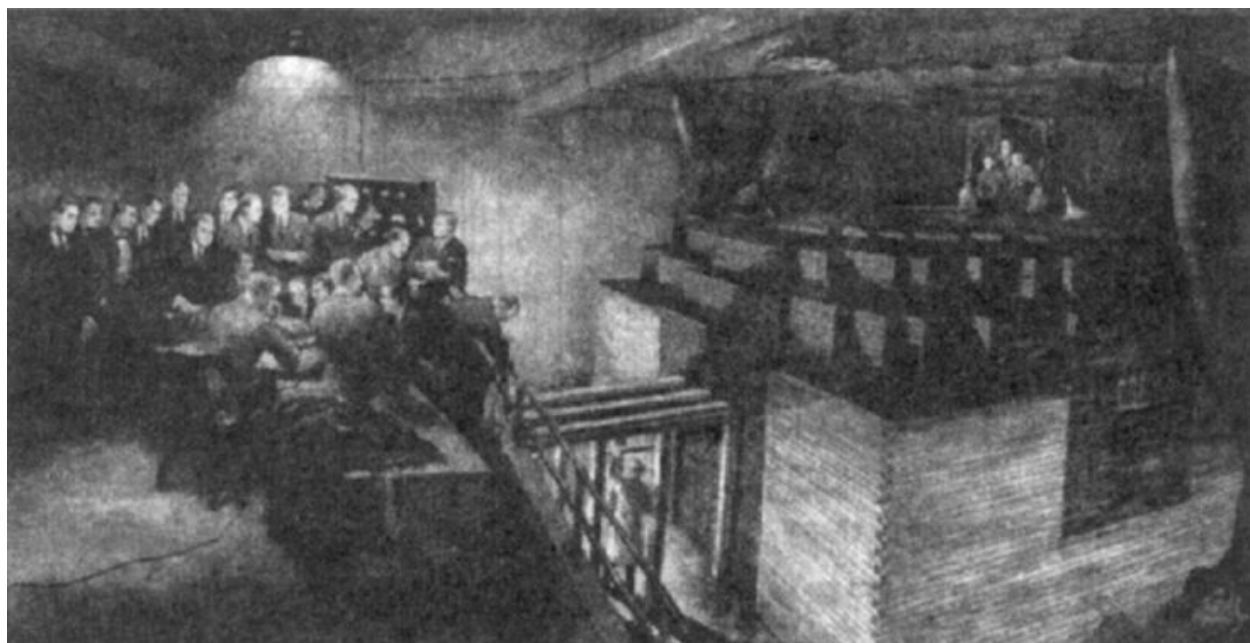


图8-11 第一座核反应堆

核电的发展

原子弹是快中子引爆，快中子诱发核裂变的概率低，所以铀必须浓缩。而反应堆个头可以大，用慢化剂可以把中子速度减慢，慢中子容易被铀核吸收，裂变概率高，所以对铀的浓度要求不高。图8-12是裂变反应堆的示意图。第一座核电站是苏联造的。现在核电在法国占80%左右，在美国占20%，其他很多国家也都有很多，中国核电占的比例并不高。虽然说核电不够安全，但是现在来说大致上还是能控制的。如果你不用核电，只能用石油、煤什么的。但石油、煤是会烧光的。烧石油很可惜，石油是很好的化工原料。有人说烧点也不算浪费，能取暖啊。门

捷列夫说了一句很有名的话：“你要知道钞票也是可以用来生火取暖的。”

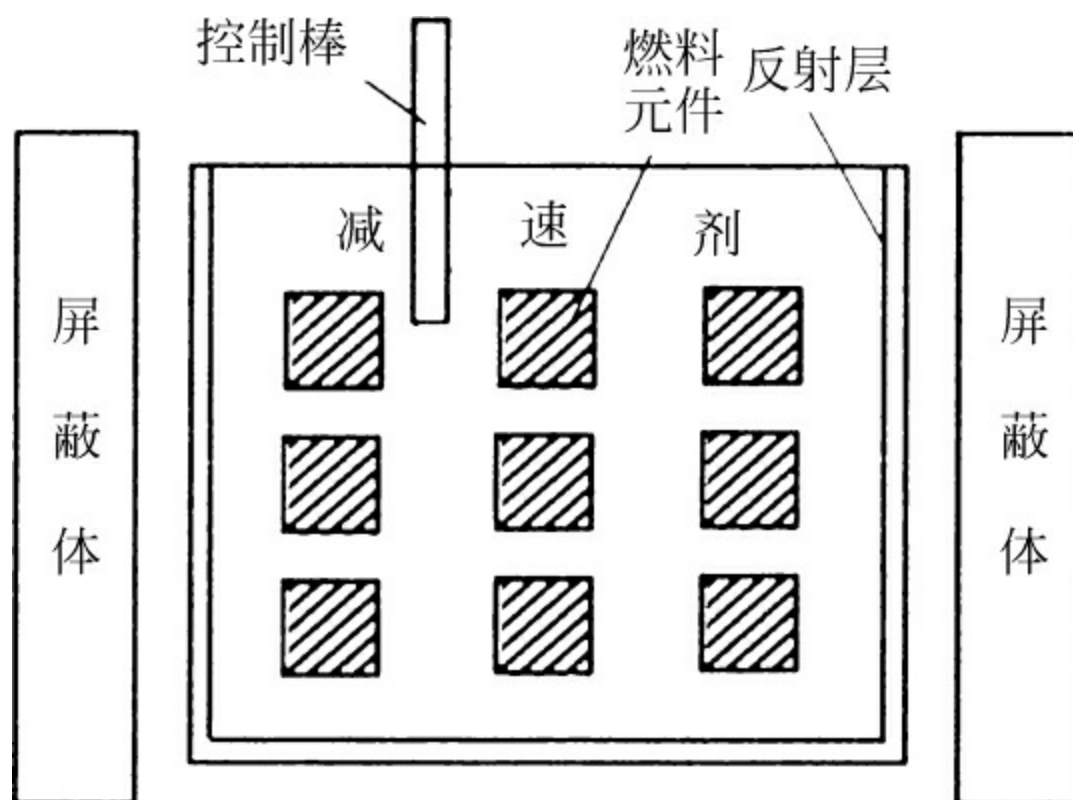


图8-12 裂变反应堆示意图

取之不尽的聚变能

不过裂变的燃料铀等元素是有限的，在地球上并不多，用不了多少年也会用完。真正可以长期利用的是聚变的原料，这是取之不尽用之不竭的。1升海水中提炼出来的氘用于聚变，相当于300升汽油。对人类来说这是用不完的。问题是没有很好的方法控制热核反应，控制聚变反应。人类知道的唯一的办法就是做成氢弹，咣一下炸了，别的办法都没有。

现在就是要想办法来控制热核反应。另外，用氦3作原料比用氘和氚要好。用氘和氚的话，产物当中有中子，中子穿透力太强，对设备的毁坏很厉害，但是用氦3呢，它的生成物中没有中子，生成的是质子。

质子带电，容易屏蔽。月亮上有大量的氦3，所以登月这件事情，从长远看也有实用意义。有人算过，要是运回一飞船的氦3的话，就够人类使用一年。但是聚变反应控制起来很困难，因为聚变反应需要上亿度的高温。反应的“容器”用什么材料制造，如何使反应发生，又如何控制，如何把能量引出来，都是难题。现在是试图用托卡马克（超强磁场形成的容器）或激光束聚焦，来实现可控热核反应。各个国家都在研究，包括中国也有这类装置。但是，离实用还有很长一段距离。

第八讲附录 汤川对介子质量的估计

在原子核的狭小空间里，聚集着大量的中子与质子，由于核内的质子靠得很近，相互间一定会有很强的正电排斥力。所以科学家们推测，原子核内的核子（质子与中子统称核子）之间应该还存在一种极强的吸引力，能够克服正电荷间的排斥力，科学家们称其为核力。核力应该是一种短程力，只在原子核大小的尺度（ 10^{-15} 米）下存在，超出原子核外就迅速消失。由于它比电磁力强得多，人们称之为强相互作用（强力）。1935年，日本物理学家汤川秀树提出交换力的思想来解释核力，并因此获得了1949年诺贝尔物理奖。他认为质子与中子等核子之间的核力，是由于交换某种粒子而产生的，汤川还利用测不准关系，预言了这种粒子的质量。

汤川是这样估算传播核力的粒子的质量的：核子间交换粒子的过程能量不守恒，因此只能是测不准关系允许的虚过程，交换的粒子只能是虚粒子，虚粒子可视为以光速 c 传播。核子间的距离大约是原子核半径 r 的大小，因此交换虚粒子的时间大约是

$$\Delta t = \frac{r}{c} \quad (8.4)$$

从式（8.2）所示的测不准关系

$$\Delta t \Delta E \sim h$$

可估算出被交换粒子的能量为

$$\Delta E \sim \frac{h}{\Delta t} = \frac{hc}{r} \quad (8.5)$$

再从质能关系式（8.3）

$$\Delta E = \Delta m \cdot c^2$$

可知，这种粒子实化后的质量大约为

$$\Delta m = \frac{\Delta E}{c^2} \approx \frac{h}{cr} \approx 200m_e \quad (8.6)$$

式中 m_e 为电子质量。由于这种粒子的质量介于核子和电子之间，汤川称其为介子。

汤川的理论被物理界普遍接受，但是，这种力的规律究竟如何用数学公式来描述，当时在理论上还不清楚。

今天，强相互作用理论已经取得了长足的进展，量子色动力学给出了强作用的严格的数学表达式。新理论继承和发扬了汤川的“交换力”的思想。另外，汤川估计介子质量的方法，对后人也有启发作用。

汤川秀树是最早对近代自然科学做出重大贡献的亚洲人之一，他是日本民族的骄傲。在那个种族主义甚嚣尘上的时代，他也为所有黄种人树立种族自信心做出了贡献。

第九讲 漫步太阳系

现在介绍一下太阳系。主要分四个部分讲：太阳与月球；行星与卫星；小行星与彗星；天文仪器。

1. 太阳与月球

图9-1是太阳系的示意图，中间是太阳，然后是一颗一颗的行星，外边这颗是土星，你一看光环就知道是土星。还有彗星，走的是很扁的椭圆轨道。其实有的彗星走的轨道是抛物线，还有的是双曲线，走这两种轨道的彗星一去就不再返回了。我们最注意的是那批走椭圆轨道的，他们离去后还会再回来。



图9-1 太阳系示意图

太阳简介

图9-2是太阳。太阳表面温度是6000度，中心温度是1500万度，那里不断地进行着氢聚合成氦的热核反应，维持它的生存。太阳属于主序星，处于恒星的中青年时代。太阳在这个时期能维持100亿年，现在过了50亿年，还有50亿年基本上会是现在的样子，所以我们可以放心地活着，没有问题。太阳表面有很多耀斑、黑子（图9-3），黑子是太阳

表面的旋风。黑子的温度都是几千度，只不过比其周围的温度稍微低一点，所以你觉得它好像处于低温。其实不是，也是高温，只不过外面6000度，它不到6000度就是了。

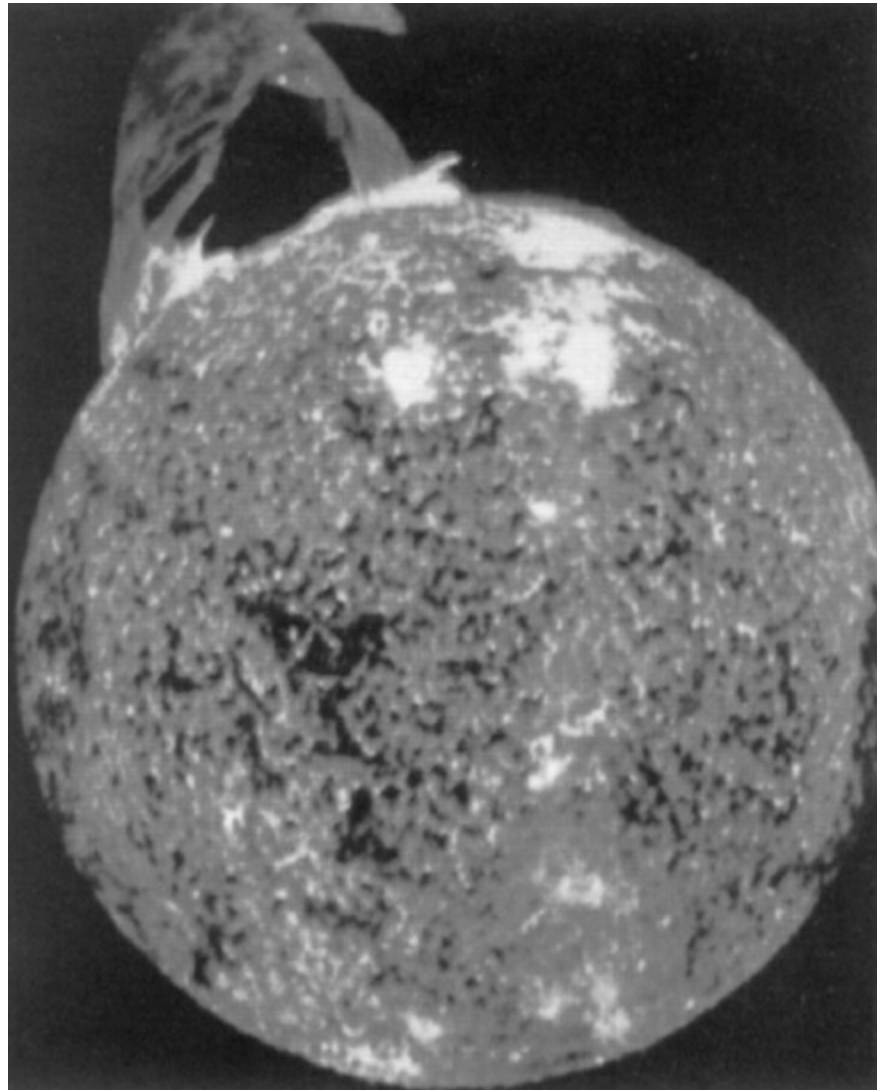


图9-2 太阳

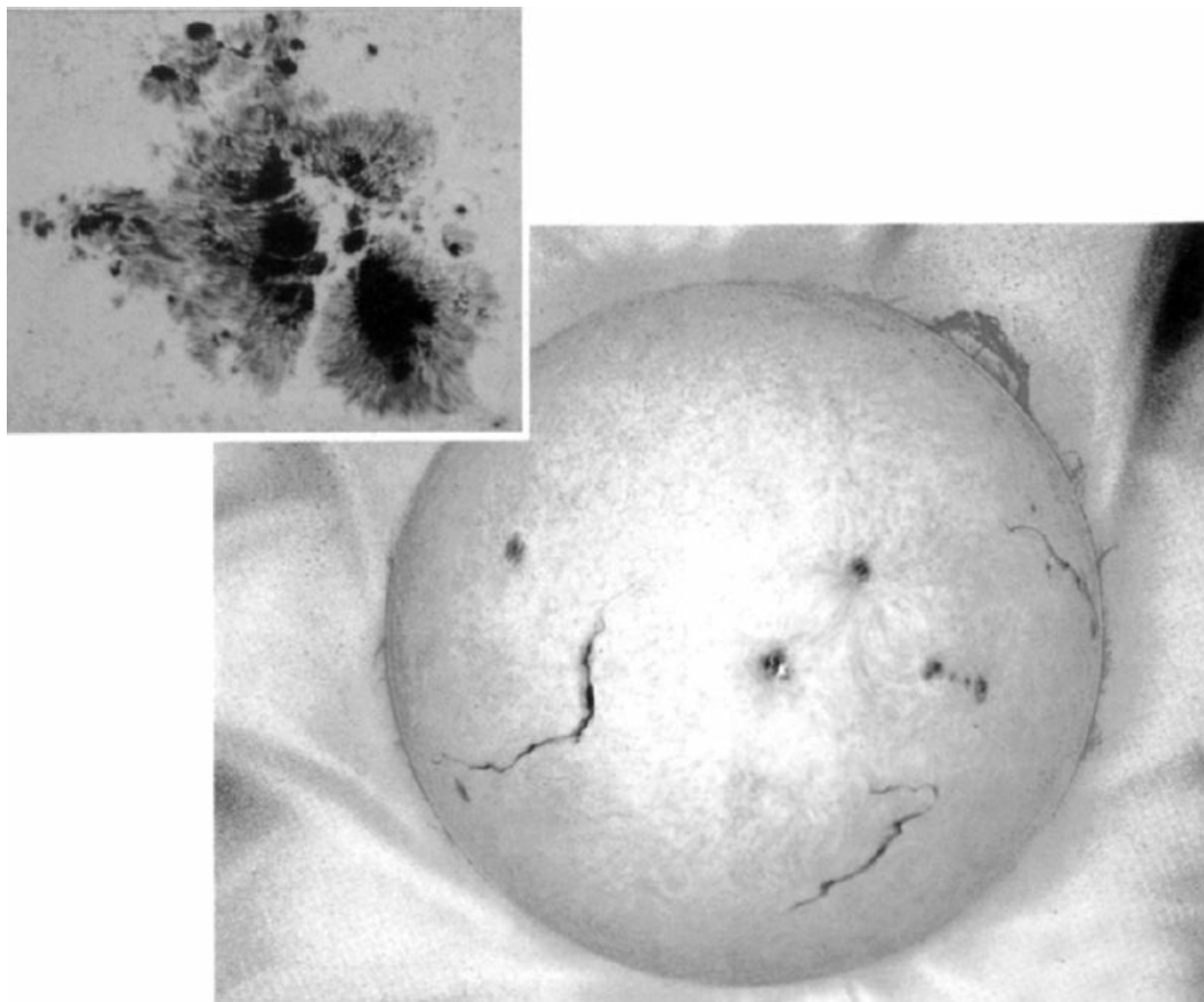


图9-3 太阳上的黑子

太阳的质量大概占太阳系总质量的98%~99%。你看那一个一个的黑子，都是可以把地球放进去的。太阳的质量相当于33万个地球。对于太阳我就简单说这些。你们要是感兴趣，还可以在别的书籍中看到，我就不多说了。我说一些我觉得比较有意思，你们也可能觉得有意思的东西。

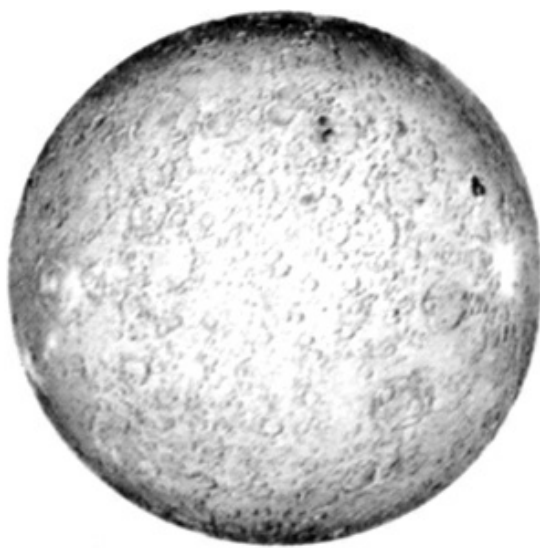
月球与探月历程

图9-4是月亮的照片。我们人类在发射月球火箭之前，只看到过月球的正面，也就是左面这一张照片。因为月亮一直是用同一面对着我们的，它围着地球转的公转角速度和自转角速度是一样的，所以老是用这

张脸对着我们，我们看到的就老是这个面。



月面照片



月球背面（在地球上看不见）

图9-4 月球的正面与背面

一直到1959年苏联发射月球火箭，它连着发射了三枚，当然也还是隔了一些时间，但隔得不是很长。第一枚从旁边过去了，第二枚直接命中月球，第三枚围着月球转，然后拍回了月球背后的照片，这样我们才看到月球的背面是什么样的。在没有看到月球背面的时候，科幻小说就可以随便地想象，例如说：月亮都快裂开了，背后有个大口子。反正谁也看不见，科幻小说可以这样写。

从图9-4中大家可以看到一个一个的环形山。环形山的起因在历史上曾有过争议，有两种观点。一种观点认为它是陨石撞的；另一种观点认为这是火山爆发留下的。

主张火山爆发的人说，你仔细看这些环形山，这些环形山中间往往有一个尖，那就是火山口。主张撞击的人就做了一个实验，弄了一摊稀泥搁哪儿，拿块石头往里一扔，“咣当”一砸，一弹，弹起一个尖来。所以那个尖不一定是火山口，不足以说明环形山是火山爆发形成的。

后来的研究表明，环形山基本上都是撞击的结果。这两张图看得还

不是很清楚，一会儿我们看水星，就会看得很清楚。

大家再看图9-4这两张月面图，月亮冲着我们的这一面是比较平的，背对我们这一面，跟一个麻子似的，坑坑洼洼的。为什么呢？因为背对我们的这一面，被撞击的几率太多了。撞地球的很多天体都撞在它上边了。从地球这边撞过去的天体，则被地球挡住，撞在地球上了。所以月亮冲着我们的这一面显得是比较平的。

图9-5，图9-6是登月的图片，载人登月是美国完成的。本来美国一心认为航天事业肯定是他们领先，结果没想到苏联捷足先登，先发射人造卫星，然后载人宇宙飞船，然后就是月球火箭。美国人赶紧急起直追，后来美国首先完成了载人登月。登月确实非常了不起。图9-5是人类踩在月球上的第一个脚印。这个宇航员在登上月球的时候就说：我迈出的是一小步，但是对于整个人类来说这是一大步。他说得完全正确。



图9-5 人类在月面上的第一个脚印



图9-6 登月的宇航员

当时，三个宇航员坐着飞船去，围着月亮转的时候两个人下来了，一个人没下来，留在上面看守那艘飞船。登月舱降落之后，一个人小心翼翼地走下来，先踩了踩底下，踩实了然后再走。因为当时很怕底下是松的，如果咕咚一下陷下去了，那就麻烦了。结果还好。美国的登月有一次很惊险，他们有一艘阿波罗13飞船，13这数字确实不大吉利，结果飞上去之后，飞船发生了故障，既去不了月球，也回不了地球，就悬在空中。然后美国总统带头向上帝祈祷，最后终于还是回来了。

宇宙航行绝对是有风险的，所以咱们发射飞船也会有风险。

月亮表面没有空气，也没有水。月球南极的中心，有人认为那里可能有冰，但还没有证实。

2. 行星与卫星

地球：我们的家园

我们再来看一看这八颗行星。图9-7是地球。这张照片是在月亮的上空拍的地球照片。底下大的部分是月亮的表面，空中悬着的是我们的地球。



图9-7 从月亮上看地球

其实地球外面的空间充满了各种各样的电磁场（图9-8）。来自太阳的粒子流形成太阳风。这些喷射出的粒子流是带电的，所以它会对地

球的磁场产生影响。这一内容是空间科学研究的重点。

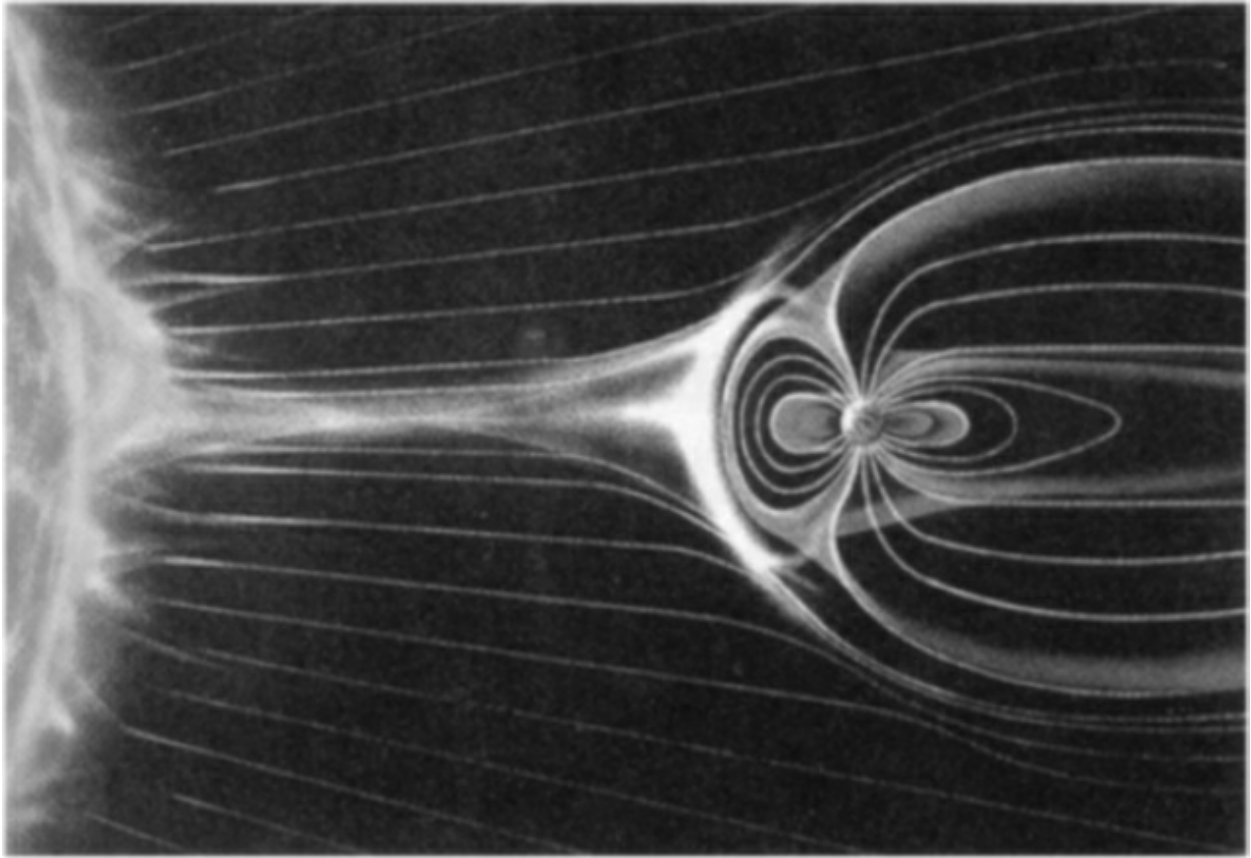


图9-8 太阳风与地磁场

水星：离太阳最近的行星

图9-9是水星。假如不认得月亮的人一看，觉得这东西好像是月亮。水星跟月亮很像，也是没有水没有空气，上面有大量的环形山，环形山中间的尖很明显。这些环形山都是撞击的结果（图9-10）。

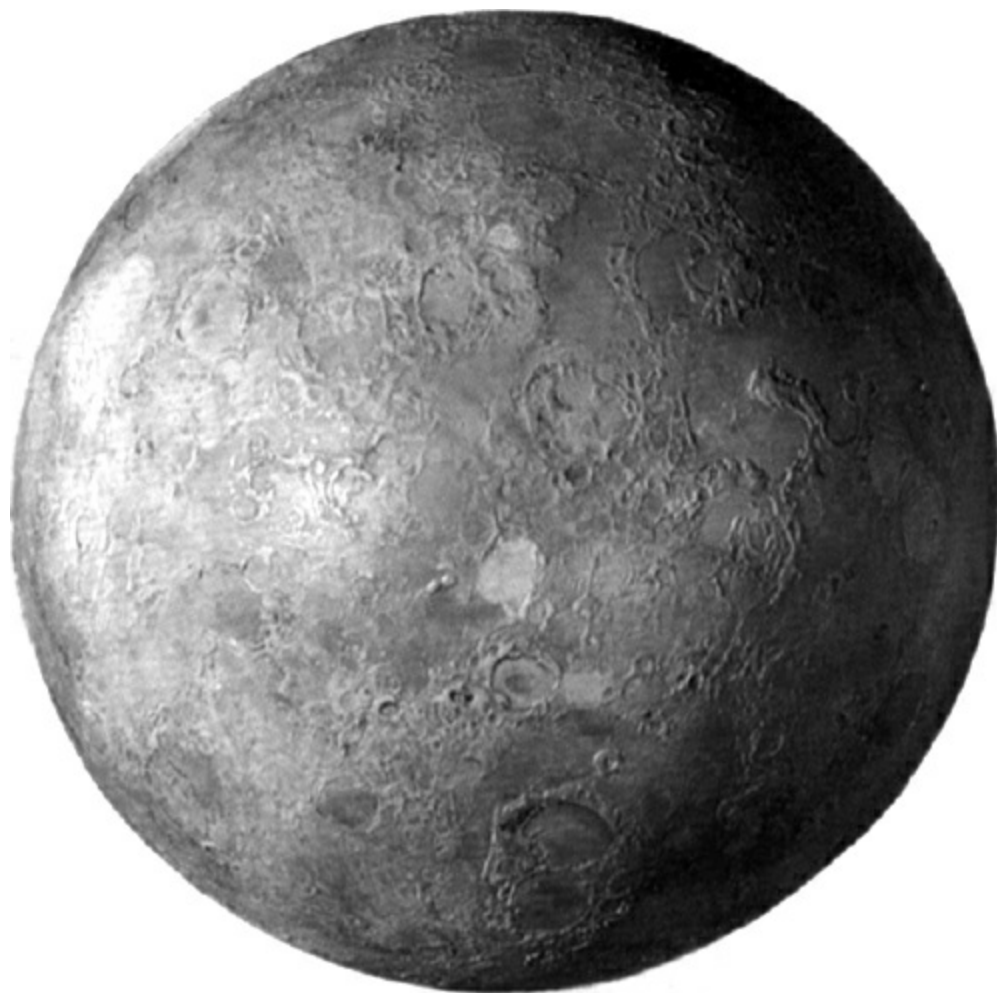


图9-9 水星



图9-10 水星的表面

地球的邻居：金星与火星

我们来看一下金星（图9-11）。我现在讲的顺序是按照各颗行星离太阳的远近，从近到远来讲。离太阳最近的行星是水星，然后是金星。金星是我们肉眼看到的天空中最明亮的一颗行星。这颗星早晨和晚上出现，非常明亮。



图9-11 金星

人类其实看到的都是金星表面的大气，有很浓厚的云把它盖住，但是人类一开始不知道，后来望远镜的技术好了，再加上有了其他的探测手段之后，人们才知道我们看到的并不是金星的固体表面，看到的只是外面一层很浓厚的云。

因为金星和火星是离地球最近的两颗行星。火星比地球离太阳稍微远一点，而金星则稍微近一点，所以这两颗星，引起人类更多注意。原因之一是推测它们与地球的状况可能相近，也许会有高级生命存在。历史上曾有人猜测是否有火星人和金星人。人类首先注意的是火星，因为火星看得比较清楚。

消逝的火星人

我们先看一下火星。火星的大气比较稀薄，它表面是红色的，南北两极都有白色的东西，而且夏天的时候这白色的极冠会缩小，而冬季的时候极冠会加大（图9-12）。

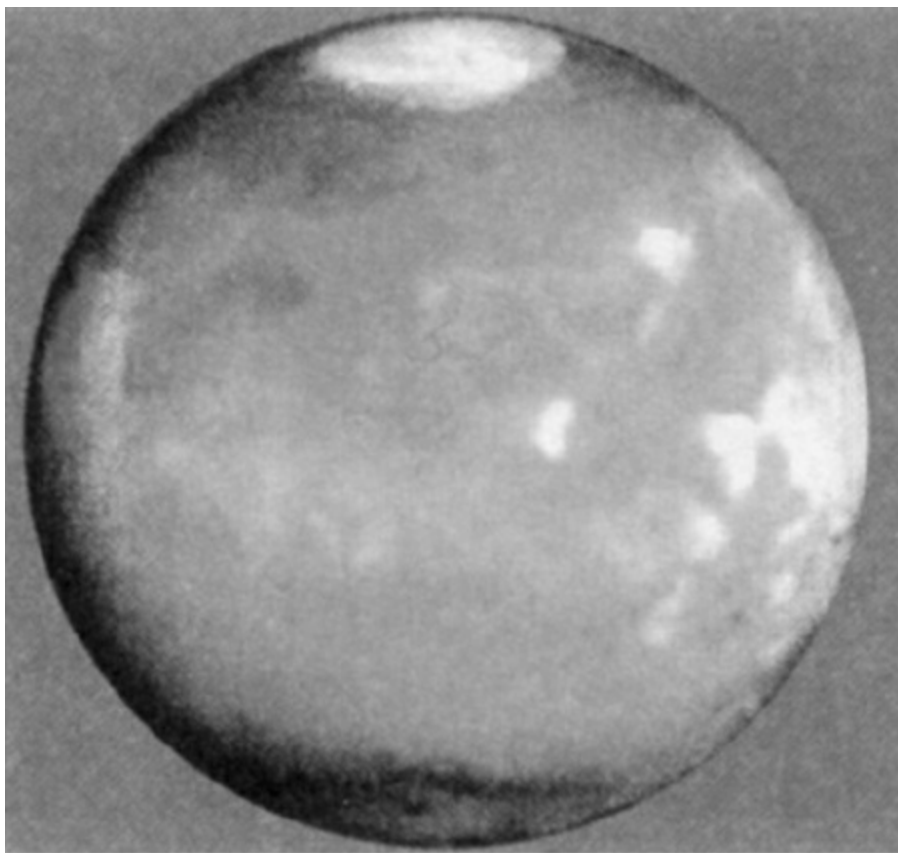


图9-12 火星极冠

人们以为那些白色的东西是冰雪。当时望远镜不太好，感觉在火星表面似乎有很多条纹。有一些东西走向的，还有一些南北走向的。最初以为是火星人修的运河。有些天文学家认为火星上这么红，肯定比较干燥。看来火星人还是很发达的，他们用两极融化的雪水来灌溉。后来望远镜比较好的时候就看清楚了，那些“线条”不是运河，只不过是火星上的地貌。那是很多小黑点，只不过你看不清，误看地连在一起了。

人们在看东西看不清楚的时候，往往你觉得它像什么，你看它就越

像什么。后来发现火星的条件比较恶劣，不像我们人类原来想象的那么好。大概有点像南极洲的那种温度，但是大气要稀薄得多，有高级生命的可能性很小。

火星探测器发射以后，大家看清了，火星的表面，就跟戈壁滩一样（图9-13）。但是人类觉得，火星的表面似乎有被水、被液态的东西冲刷过的迹象，还有很多人寄予希望，在火星的表面底下是不是有大量的水。



图9-13 探测器在火星表面

如果有水，就可能有生物，而且就可能有比较高级的生物，比病毒要高级一些的。因为宇宙当中肯定有很多病毒，这种最低等的生物肯定广泛存在。但人类感兴趣的不是那些东西，而是比较高级一点的，特别是有没有外星人的问题。

此外，火星有两颗卫星，刚开始看不太清楚，发现火星的卫星个儿都不大，有人就猜测它们是不是人造卫星，是不是火星发射的人造卫星。看清以后，其实就是两块大石头，比较小，形状不规则（图9-14）。

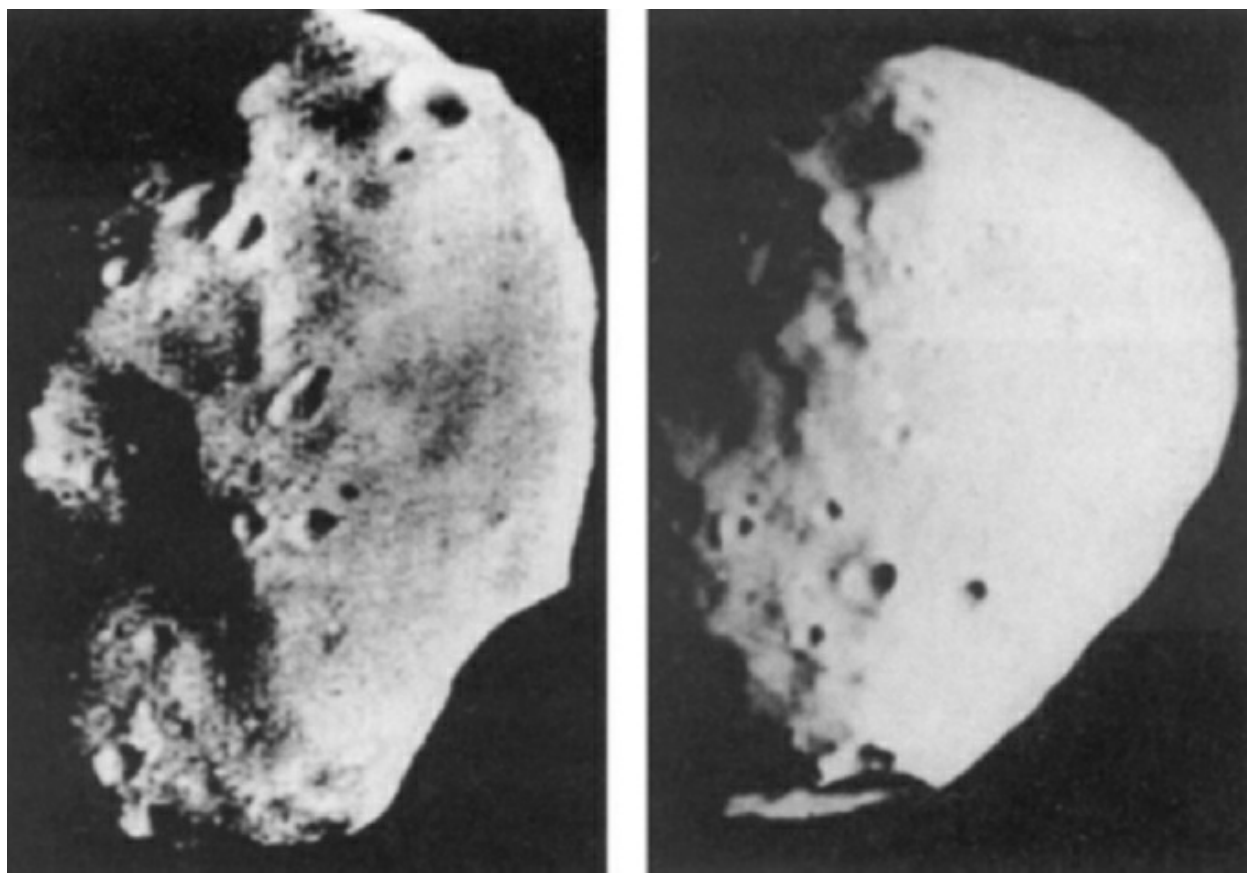


图9-14 火卫1与火卫2

我有一讲曾经讲过，开普勒曾猜测火星有两颗卫星。他的理由是，地球有一颗卫星，木星当时看到四颗卫星，火星位于地球和木星之间，他认为应该有两颗卫星。他认为上帝在创造宇宙的时候肯定有想法，应该有个规律，上帝一定不会乱造。他推测火星卫星应该有两颗。结果，他还真说对了，火星的卫星还真是两颗。

金星：大气的高压与高温

后来人们认识到，火星上的自然条件比地球恶劣得多，火星肯定是没有希望了。火星上有高级生物的可能性几乎为零，不会有植物、动物。于是，人们就把注意力集中到金星。那时候已经知道金星表面有一层很浓厚的云，看不见金星的地表。而且金星的大气很浓，主要成分是二氧化碳。主要是二氧化碳倒没有关系，地球历史上也曾经有过大气成

分主要是二氧化碳的时期，这不是什么大问题。

后来，苏联发射了金星探测器，在上边实现了软着陆。探测器落下去以后才发现，那里条件特别恶劣。金星的表面有90个大气压，480度高温，根本没有液态的水，当然在大气当中还是有水的成分。而且探测器在降落的时候，还要穿过一层浓硫酸构成的云，所以金星上面有生物的可能性微乎其微。大家很失望，看来金星也没有高级的生命。

美国的金星探测器上去了很多，咱们中国还没有，我想，咱们大概不久之后会发射的。因为现在我们中国的科学技术正在发达起来，而且又有钱，对吧！所以我们就有可能进行更多的空间探测。希望大家要努力学习，争取也去参加这样一些工作。我们有一些毕业的同学就在空间中心这样的地方工作。

图9-15是金星上的山，这是探测器在金星上降落以后，在金星表面拍摄的。金星和火星，我就讲这么多。



图9-15 金星的地表

木星：“木纹”与“大红斑”

下边我们再来看一下木星。伽利略使用望远镜的时候，就观测到在木星上边有“木纹”（意为木星上的条纹）。木星图像（图9-16）中一根一根的横纹，还真的有点像木头纹似的，这是一种巧合。另外，木星左下部有一块大红斑，从伽利略时代就发现的这一块大红斑，一直保留到现在。



图9-16 木星

大家后来知道，木星不是一颗固体星，而是一颗流体星。外边的大气，主要由氢组成，还有一些氦。大气下面，有氢组成的海洋，大概有五六万公里深。都是液态的分子氢或者金属氢构成的。中心有铁和硅构成的固体核，跟海洋比，体积就小了不少。木星主要是一颗流体星。人们还发现了木星的几颗卫星。那几颗大的卫星，伽利略的时候就发现了，所以叫做伽利略卫星。

木星既然是一颗流体星，为什么会有一个大红斑老是在那个位置上不动呢？以后人们才知道，那是一个旋风。木星上的大旋风。这个大旋风已经存在几百年了。这个斑为什么是红色的呢？因为含有大量的磷的化合物。这个红斑很大，可以把地球搁进去。

木星的卫星：是否存在生命？

木星的卫星，我们现在兴趣比较大。这些卫星，都是一些固体星。这些固体星，有的表面底下可能有水。因为有些探测器从木星旁边飞过的时候，收集过一些数据，觉得下面可能有液态的水，而且，这些液态水可能有盐分，这样就有了存在生命的可能性。当然，有特别高级生命的可能性并不大。

图9-17是木卫一，图9-18是木卫二上的冰缝和陨石坑，图9-19和图9-20是木卫三和木卫四的表面。未来大概会加快对木星卫星的探测，因为那些地方应该是可以降落的。木星表面都是流体，没法儿降落，只可以围绕它转。不过木星的这些固体卫星是有可能降落的。



图9-17 木卫一

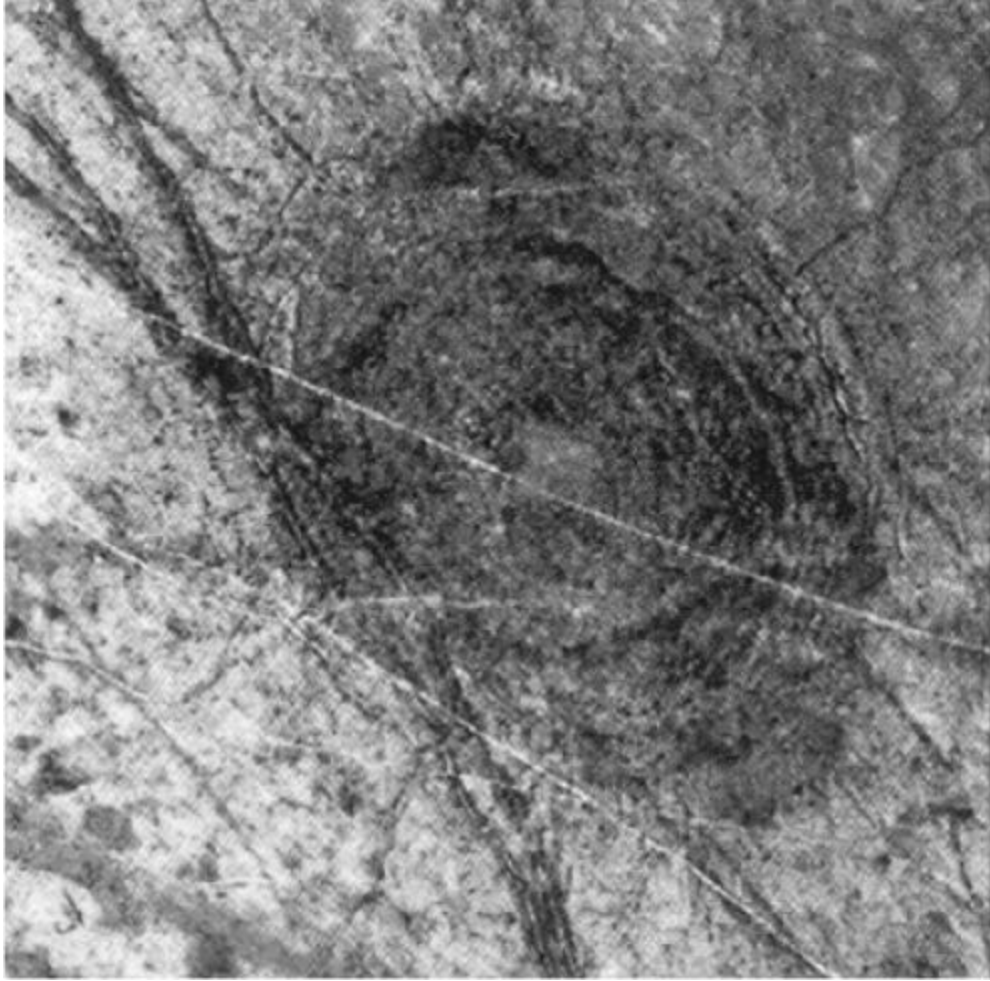


图9-18 木卫二上的冰缝和陨石坑

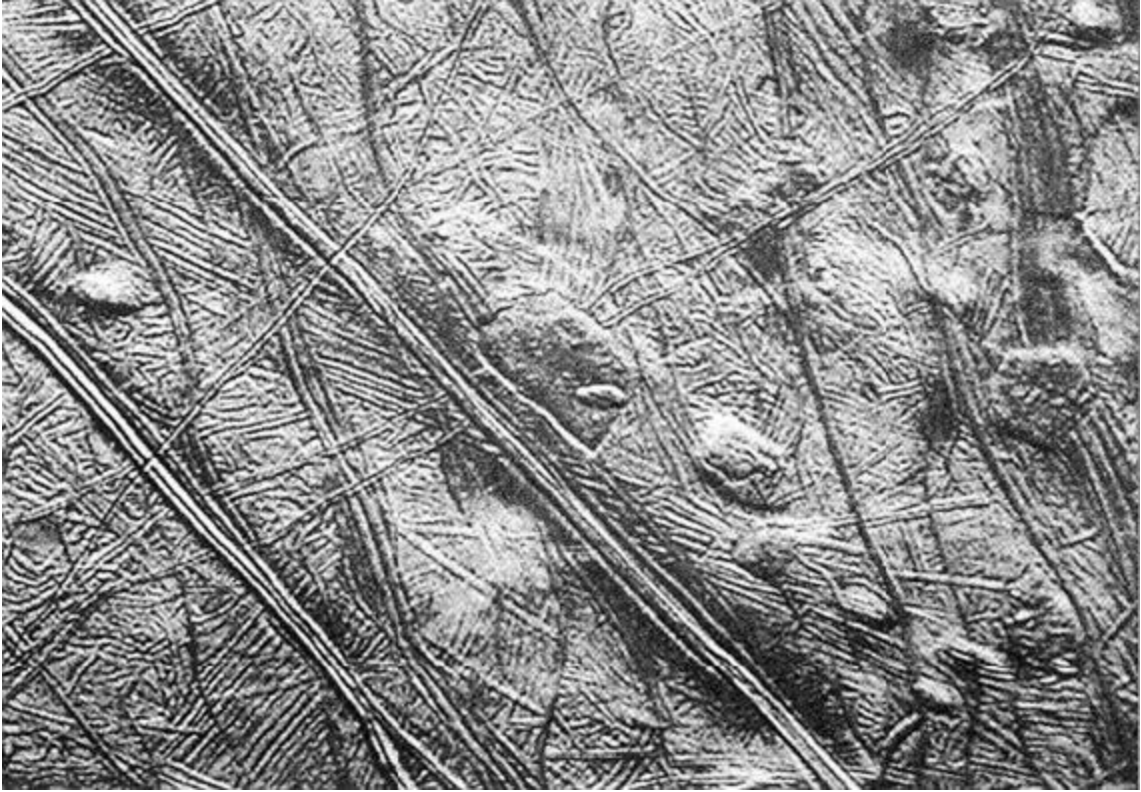


图9-19 木卫三的局部表面



图9-20 木卫四的地貌

近年来，大家对木星有一个怀疑，天文观测发现，木星放出去的热量比它吸收的热量要多，所以有人怀疑木星实际上是一颗恒星。它要真是一颗恒星，太阳和木星就构成一个双星系。但是这个观点并没有引起太大的响应，还需要进一步研究。

土星和它的光环

我们再看一下土星（图9-21）。土星特别引人注目的是它的光环（图9-22）。这个光环是伽利略首先发现的。但是伽利略没有认出这是光环，只知道土星有附属物。

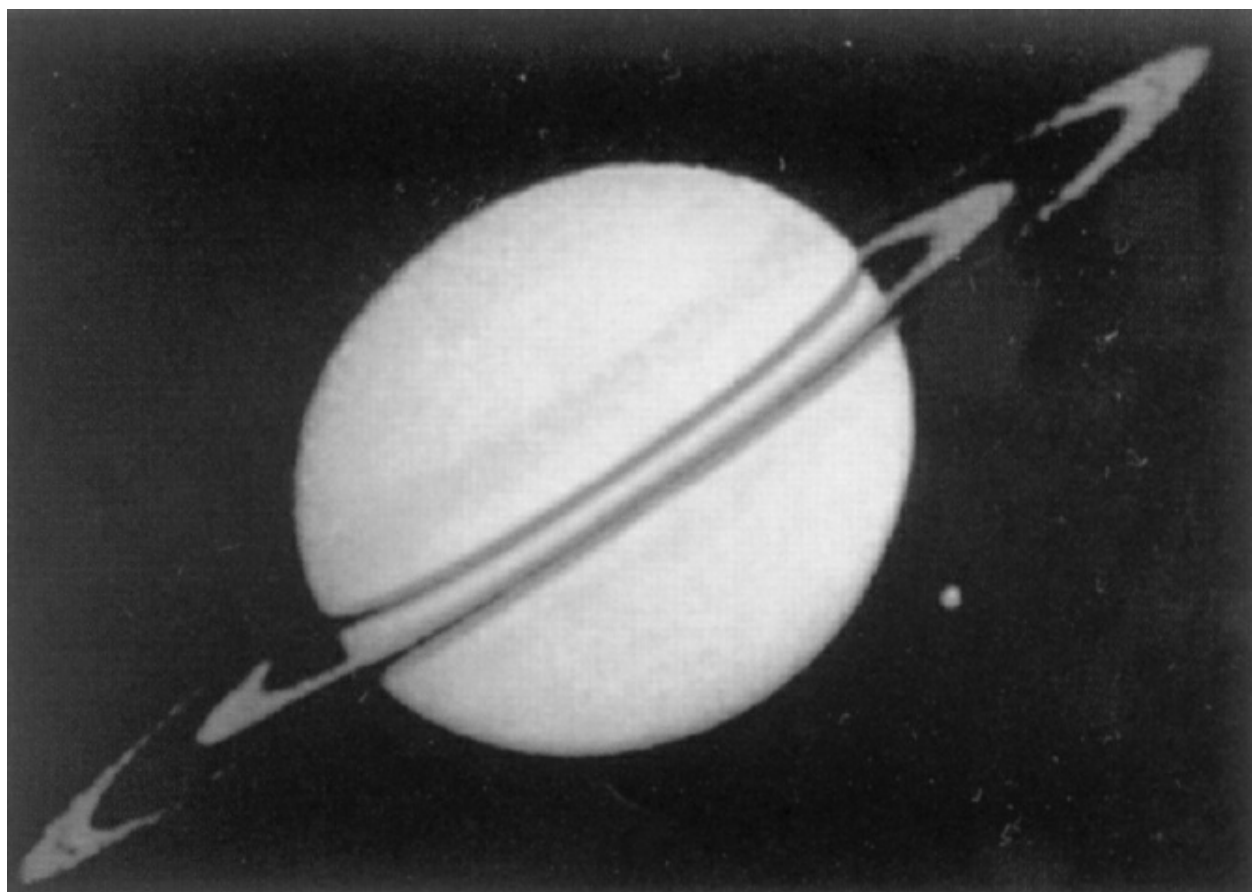


图9-21 土星

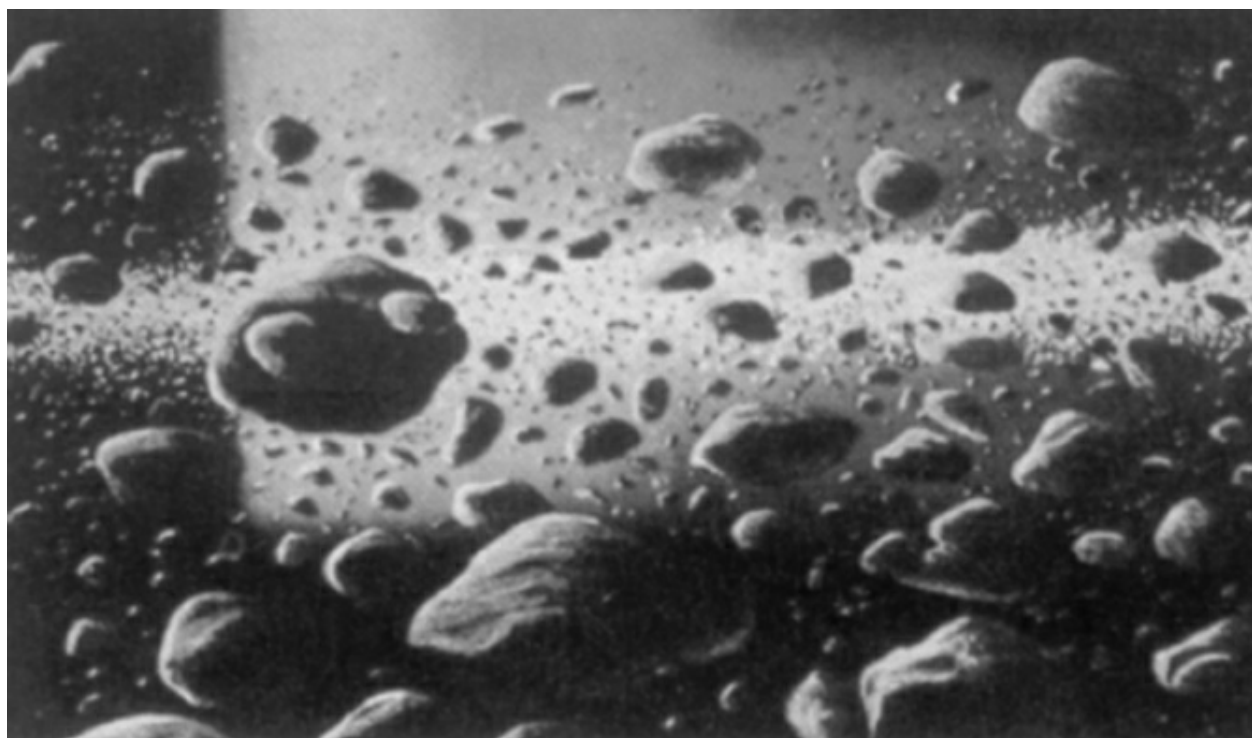


图9-22 土星光环的组成

后来有一个叫休琴士的人认出来了。休琴士看出这个附属物是光环的时候，非常高兴，编了一个密语。过了三年，确认是光环以后就公布了密语。过了一段时间之后，这个光环没有了。

为什么光环会没了？因为这个光环很薄，当盘面朝向我们的时候，我们就可以看见光环。当盘面侧过去，完全从侧面看，很薄很薄的一层，望远镜不太好的时候就看不清楚，所以光环又没了。

那个时候报纸上就开始登，说光环碎了，碎片正在飞向地球。耸人听闻，这样报纸可以卖得快一点。后来，角度又转过来了，又看见光环了。流言便不攻自破了。

光环是什么东西呢？主要是由冰块、石头块组成的东西。因为这些碎块太多了，所以从远处看像光环。

天王星：用望远镜发现的第一颗行星

在土星之外是天王星。天王星是用望远镜发现的。肉眼看见的就是金、木、水、火、土五颗行星。金、木、水、火、土这五颗星里，比较难以看见的是水星；我们最容易认出的是金星，一般人都认得，早晨叫启明星，黄昏叫长庚星，都是它；火星有点发红，在天上也容易看出来；木星和土星也能看到，它俩颜色差不多，一般不是很熟悉天文的人不大容易分得清。

水星是很难看得到的，它离太阳太近了。据说哥白尼一辈子都没见过水星。主要是因为东欧这个地方，老是阴天，老是有云彩，特别是在黎明和黄昏的时候一般都有云，看到水星的可能性就很小。

有了望远镜之后，人们发现了天王星。图9-23是用哈勃望远镜拍摄的天王星照片。天王星很有意思，它的自转轨道和公转轨道是垂直的。也就是说它的自转轴是沿着它的公转轨道面的，结果它就躺在公转的轨道上了。

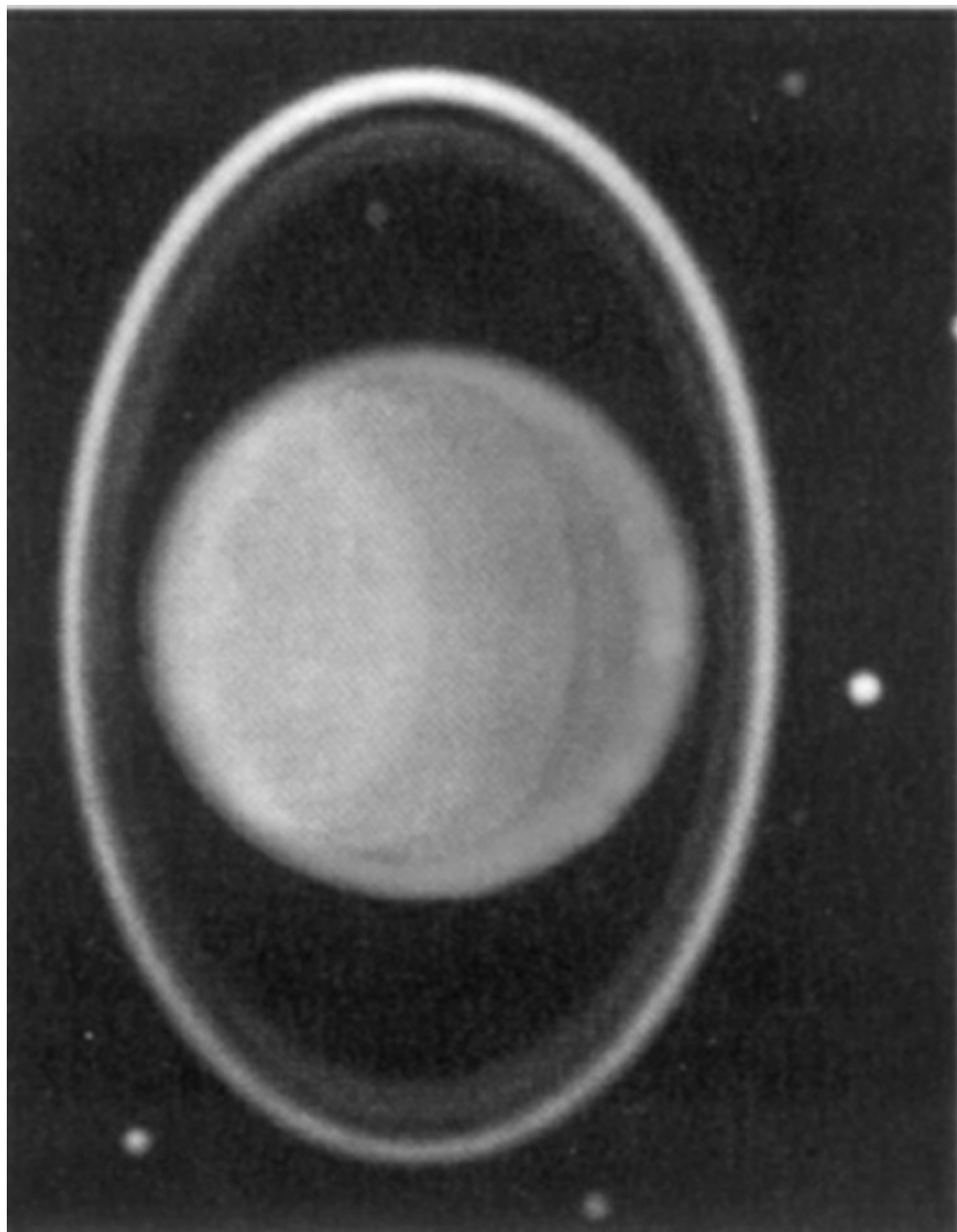


图9-23 天王星

海王星：万有引力定律预言的行星

下面谈一下海王星。在第二讲中我们谈到过海王星是先预言后发现的，是勒维耶和亚当斯分别根据万有引力定律预言的。

海王星发现以后，万有引力定律就得到了全面的肯定。因为一个真正的定律出来以后，不仅应该能解释已有的现象，最好还能预言新的现

象。当时已经知道万有引力可以解释哈雷彗星的轨道，这个时候又把海王星算出来了。

图9-24是从海王星的卫星上看海王星，蓝色的是海王星，这颗卫星是红色的，你看还有一个火山正在爆发，这张照片肯定不是真的，只是一张示意图，肯定是一张想象的图。

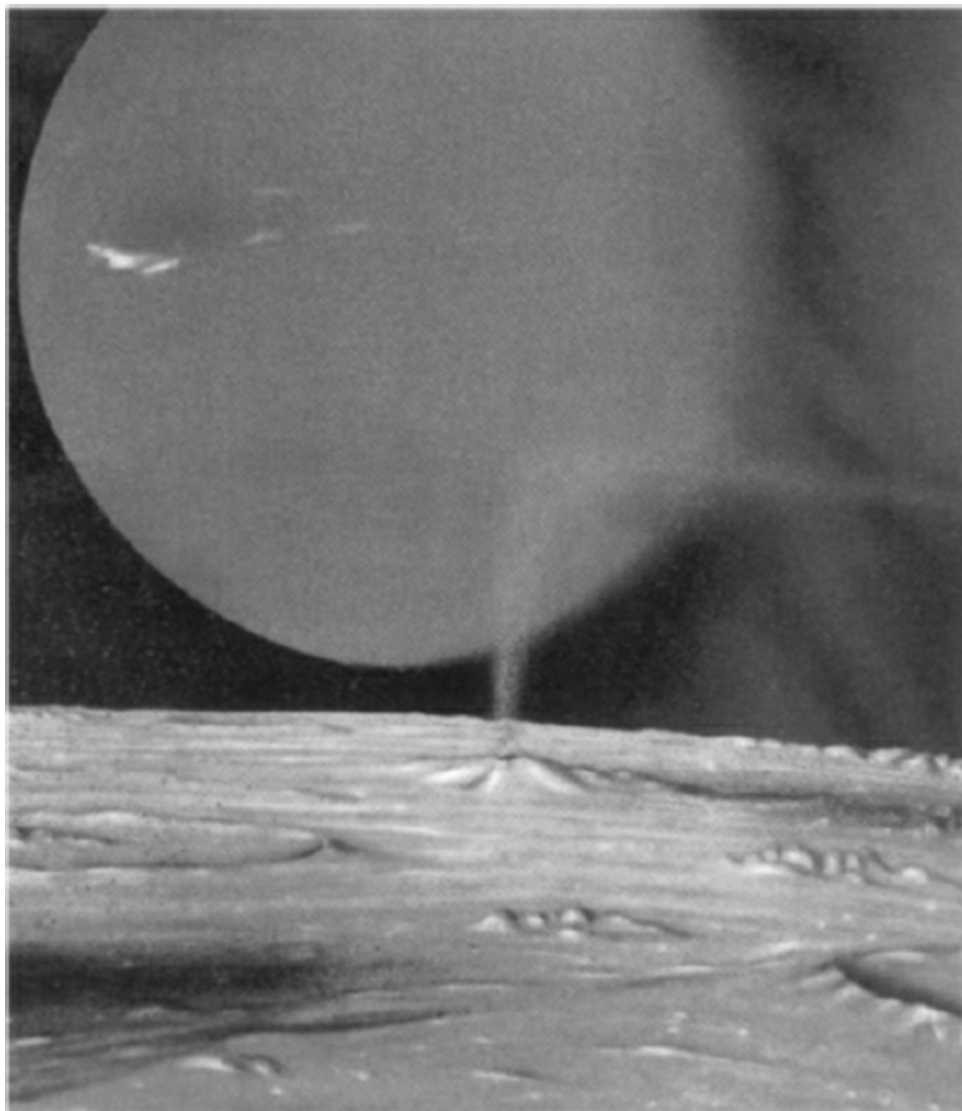


图9-24 从海王卫星上看海王星

冥王星：被除名的大行星

海王星算出来以后，又有人用类似的方法去预测更远一些的行星。

找了半天，后来找到一颗冥王星。但有人说，这个冥王星其实是偶然找到的。在预测的位置没有看见，就往旁边搜寻。找来找去，最后终于找到一个。

但是后来的研究表明像冥王星这样的星太多了，有好几个，而且有的个头比它还大，另外还有其他的原因。总之国际天文学会把冥王星的大行星资格给取消了。对这一决定，美国人很沮丧，因为他们就只发现了冥王星，而这个大行星现在又给取消了。这是因为美国的科学发展比较晚，在大行星的发现上，欧洲国家就捷足先登了。

美国这个国家是崇尚技术的，技术先发展起来，然后科学才发展起来。美国人特别愿意动手，愿意做实验；德国人则愿意思考，想一想这个问题该怎么弄。有人打了一个比方，说前面有一条路，是一个迷宫，一个美国人和一个德国人走到那儿了，美国人毫不考虑就往里走，走到前面一看，堵住了，就回来了；然后再走，堵住了，又回来；再重新走，就一直在那儿尝试。而德国人则坐在路口想，应该怎么走。两国人的风格不一样。美国的科学真正领先大概是在“二战”前后，因为那时大批欧洲的第一流学者跑到了美国。

图9-25是冥王星，现在已经被取消了大行星的资格，我想还是把它搁在这儿，大家看一看。照片中底下这个天体是冥王星的卫星，叫冥卫。悬在远方的是冥王星。这两颗星挺有意思，它俩脸对脸地转。它们的自转角速度和公转角速度都是一样的。两颗星的自转角速度一样，这两颗星的自转角速度还和公转角速度一样，所以它们俩总是脸对脸，谁都看不见谁的背面。很有意思。



图9-25 从冥王卫星上看冥王星

太阳的大家庭

我们来看看这八颗行星的轨道。现在对行星是这样称呼的。原来说说是九大行星，把冥王星资格取消以后，不就是八大行星了吗！但是国际天文学会决定不再称呼“大行星”，以后就称“行星”。行星只有两类，一类是“行星”，就这八颗；另外还有一类小天体，包括如冥王星大小的矮行星，以及为数众多的“小行星”。

大家来看，图9-26右侧是太阳和几个类地行星，类地行星质量小，密度大，都是固体星。左侧这些主要是跟木星相似的类木行星，它们质量大、密度小，其中两颗最大的（木星与土星）是流体星。类地行星的轨道分布范围很小。再往外就是木星、土星这样一些星的轨道。

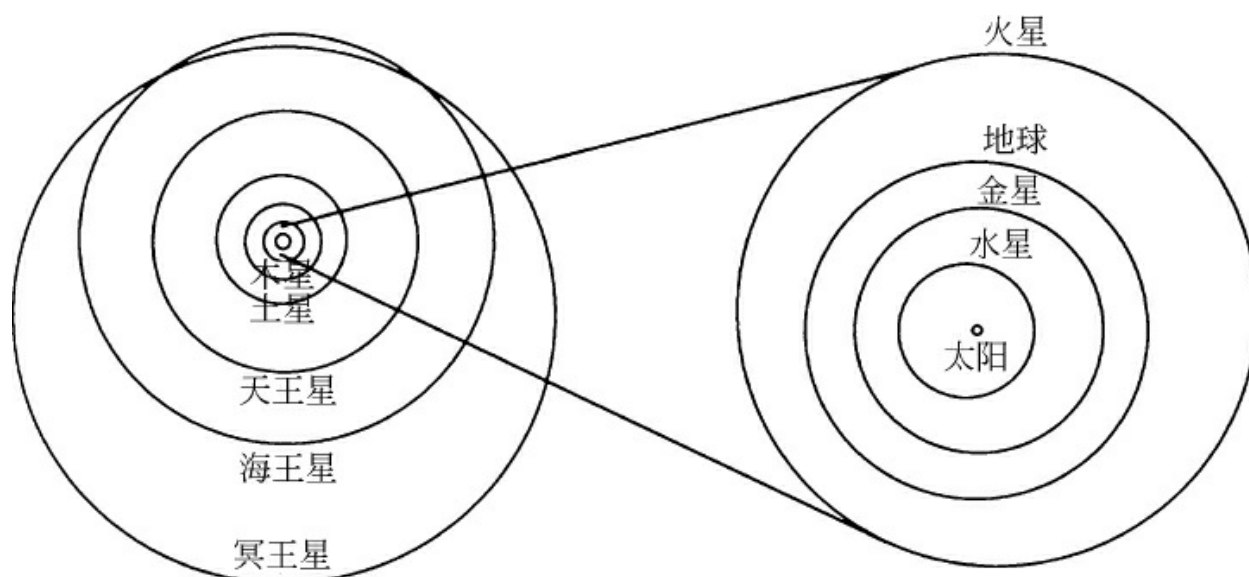


图9-26 八颗行星及冥王星的公转轨道

表9-1是八颗行星的有关数据，大家先注意，光环不是土星独有的，木星、天王星、海王星也都有，但是不明显。行星的体积是木星最大，土星其次。卫星的数目呢？水星、金星没有卫星，地球有一个，火星有两个，木星有十六个，土星有十八个。这个数据都已经很老了，现在发现木星和土星的卫星都在二三十个以上，还在不断发现新的。我们参加编一个课本的时候，有人说要用最新的材料。一查资料，木星有几个新发现的卫星，就加上去了。加上去了之后，书刚一出来，就看到报道了，说看错了，那些新发现的不是。所以紧跟也不能跟得太紧。还是要等它一段时间再说。

表9-1 八颗行星的数据

行星	日星距离 /万千米	公转周期 /地球日	赤道半径 /千米	质量 /10 ²⁴ 千克	平均密度 /(克/厘米 ³)	逃逸速度 /(千米/秒)	卫星 数目	光环
水星	5791	88	2440	0.33	5.42	4.25	0	
金星	10820	225	6052	4.87	5.25	10.36	0	
地球	14960	365	6378	5.98	5.52	11.18	1	
火星	22794	687	3397	0.64	3.94	5.02	2	
木星	77833	4333	71492	1900	1.33	59.56	16	有
土星	142940	10760	60268	569	0.69	35.49	18	有
天王星	287099	30685	25559	86.9	1.29	21.30	17	有
海王星	450430	60190	24746	102	1.64	23.50	8	有

3. 小行星

神秘的提丢斯—波特定则

现在我们来讲一下小行星的事情。18世纪德国有一个中学教师叫提丢斯，他发现，所有的行星到太阳的距离都有一个规律，什么规律呢？他给出了一个公式，叫提丢斯定则。后来柏林天文台的波特又得到了波特定则，波特定则跟提丢斯定则本质是一个，只不过两人给出的公式样子不大一样。提丢斯得到的是这么一个公式：

$$D = (n+4) / 10$$

单位是“天文单位”，地球的平均轨道半径在天文学中经常用，叫做一个天文单位。对于水星，把n=0带进去。金星把n=3带进去；然后加倍，地球n=6，火星n=12。带进去算出来的值，就是相应行星的轨道平均半径。表9-2是根据该公式算出来各行星的值与观测值。结果发现水星算出来的值0.4跟观测到的值0.39非常接近。金星算出来的0.7，也跟观测到的0.72接近。地球是规定的，是1.0；火星算出的是1.6，测量值是1.52。中间有一个n=24的地方，算出了一个2.8，但那里没有行星。n=48的地方有木星。然后有土星，还有天王星。海王星不是太准了。但是到天王星为止，都很准。天王星还没有发现的时候，这个规则就已经总

结出来了。出来以后，很多人不信，认为这东西是凑的，正好碰巧了。天王星发现以后，一看，哎哟！还真是对的。于是大家觉得不能不信。

表9-2 提丢斯—波特定则

行 星	水星	金星	地球	火星	小行星带	木星	土星	天王星	海王星
n	0	3	6	12	24	48	96	192	384
D (计算值)	0.4	0.7	1.0	1.6	2.8	5.2	10.0	19.6	38.8
D (测量值)	0.39	0.72	1.00	1.52	2.3~3.3	5.20	9.56	19.3	30.2

上帝会浪费这片空间吗

不过 $n=24$ 这儿怎么是空的啊？波特说过一句有名的话：“难道上帝会浪费这片空间吗？绝对不会。”别的天文学家觉得，上帝未必跟波特想的一样。

不久之后，在那个位置，真的发现了一颗星。命名为谷神星。但这个谷神星比月亮还小很多。不久又发现了一个，命名为智神星。智神星和谷神星加在一块儿，还比月亮小很多。

大家说，这是怎么一回事儿？有些聪明的脑袋就开始想了：这不是一个大行星碎了。如果是一个大行星碎了的话，这两颗星的椭圆轨道的交点，就应该是这个大行星碎裂的地方。所以只要把望远镜对准这两个椭圆轨道的交点，等着，这些碎片走啊走啊，一定要返回来。那么你在这儿等着，一定可以等到其他的小行星。

这首先是由一个医生提出来的。这位医生是个天文爱好者，他在病床旁边护理病人的时候，就思考这个问题。然后，晚上的时候，他就用望远镜在那儿找。等了好长时间，终于等到了一个小行星，就是婚神星。不久，人家就报导，另外一个人也找到一个，那人跟他的想法一样，但是望远镜指向的是另外一个交点，也找到了一个小行星，就是灶神星。但是这四颗星加在一起比月亮还是小很多。所以大家觉得，可能还有很多碎块。

现在我们知道，碎块确实非常多，在这个位置有一个小行星带，大概是在 $D=2.3\sim 3.3$ ，就是在 $D=2.8$ 的附近，确实是有大量的小行星存在，小行星越发现越多。

中国人的发现

我们中国人在一九二几年就发现了一颗小行星。1928年的时候，26岁的张钰哲发现了一颗小行星，发现小行星以后很高兴，他也很爱国，就以“中华号”来命名。后来这颗星找不着了，因为当时中国一直在动荡，对科学也不重视，所以他得不到什么经费。

中国解放以后，开始重视科学，但当时国家很穷，没什么钱。不过找小行星所需的成本不大，紫金山天文台在得到国家的资助后，就开始找，终于把张钰哲发现的“中华号”找到了。此后中国天文界就有一个找小行星的项目，这个项目一直维持下来。

据说中国找了几千颗小行星。现在全世界总共大概发现了二三十万颗小行星，数目还在不断增加。刚开始紫金山天文台找到的，叫紫金一号，紫金二号等等，后来就以科学家的名字命名，什么张衡啊，李政道啊，杨振宁啊，等等。还有其他很多学者的名字。

后来有一些富人愿意掏钱，掏了钱之后也可以用你的名字命名。还有一颗北师大星，那是因为咱们学校在天文界有很多校友，咱们倒不见得掏钱了，因为我们有天文系嘛！所以我们也有幸有一颗北师大星。其实就是块在天上转的大石头而已！

撞击地球的危险

当然，这种小行星要是撞上地球可不得了。你看图9-27显示的这些小行星的轨道，非常之乱，说不定什么时候就会撞上地球。要是撞上地球，影响会是很大的。图9-28是小行星撞在地球上的示意图。这一撞上，触发大规模的地震和火山爆发那是肯定的。而且大量的水汽和灰尘会飞上天空，飞上天空之后就会挡住太阳光，形成连续若干年的冬天。

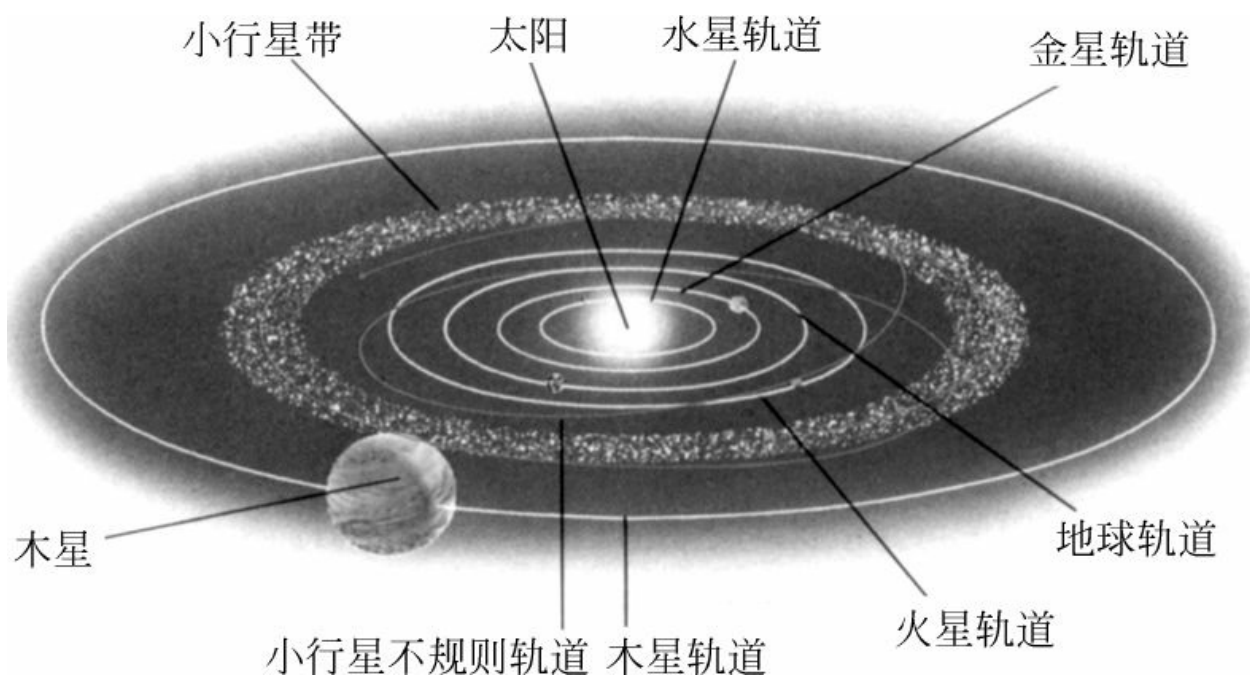


图9-27 小行星的分布



图9-28 小行星撞击地球

在研究核战争的时候，研究过核武器对人类会造成什么损害。除去直接的杀伤、放射性和冲击波外，很有可能是大量的原子弹扔下来之后，会形成核冬天，连续几年的冬天。因为灰尘和水汽都弄到天上去以后，把阳光遮住了，植物不能生长，吃的都没有了，生物必定大量灭亡。

苏美两国都研究过，如果他们打起核战争来之后会怎么样，双方都有不少的原子弹可以扔过去，除去直接杀伤外，必定会造成长期的“核冬天”，甚至可能造成人类的灭亡。所以核战争一般人也不敢轻易打。现在，比较负责任的国家，都是很谨慎的，绝不轻易使用核武器。

我们中国从一开始就宣布：我们发展核武器完全是为了自卫。并且主动承诺我们绝不首先使用核武器，而且也绝不对无核国家、无核地区使用核武器。现在中国大概是唯一一个做出这种承诺的国家，我们发展核武器是为了怕人家对我们用。你要对我们用，那我们也只好送给你了。但你要不用，无论什么情况我们都不用。我想中国这样做是非常正确的。其他国家都不承诺这一点，你要跟我打仗，说不定我什么时候就扔原子弹了。

地球上生物的灭绝，有好多种说法。其中一个说法就是小行星撞击地球。小行星撞击地球以后，会引起长时间的冬天，爬行类动物就不行了。恐龙存在的时候，哺乳类很可怜啊，我们的祖先都是哺乳类！当时都只能躲在洞里，晚上才敢出来，白天根本不敢出来。爬行类比它们凶猛多了，体积也比它们大得多。行星撞击地球形成的绵延不断的冬天，使恐龙一下子灭绝了。

当然，也有人认为是超新星爆发，射线过来了，造成恐龙灭绝。不过认为恐龙灭绝是因为小行星撞击或彗星的头部撞击地球，引起大的地震和火山爆发，造成长时间的冬天，持这种观点的人现在是多数。因为我们确实在地球上看到几个地方有大的陨石坑，比如说在美洲的一些地方，还有一些撞击的痕迹。但那都是人类文明出现之前的撞击。

4. 彗星

彗星由何而来

现在再来看一下彗星。哈雷彗星现在已经没有古时候那样大了，不那么明显了。1986年，哈雷彗星的彗尾已经开始有一些断裂，它已经比原来小多了。图9-29是1997年的一颗彗星的情况。彗星其实就是一些脏雪球（图9-30），它们就是水冰、尘埃、干冰等东西组成的一个个脏雪球。



图9-29 1997年出现的一颗彗星

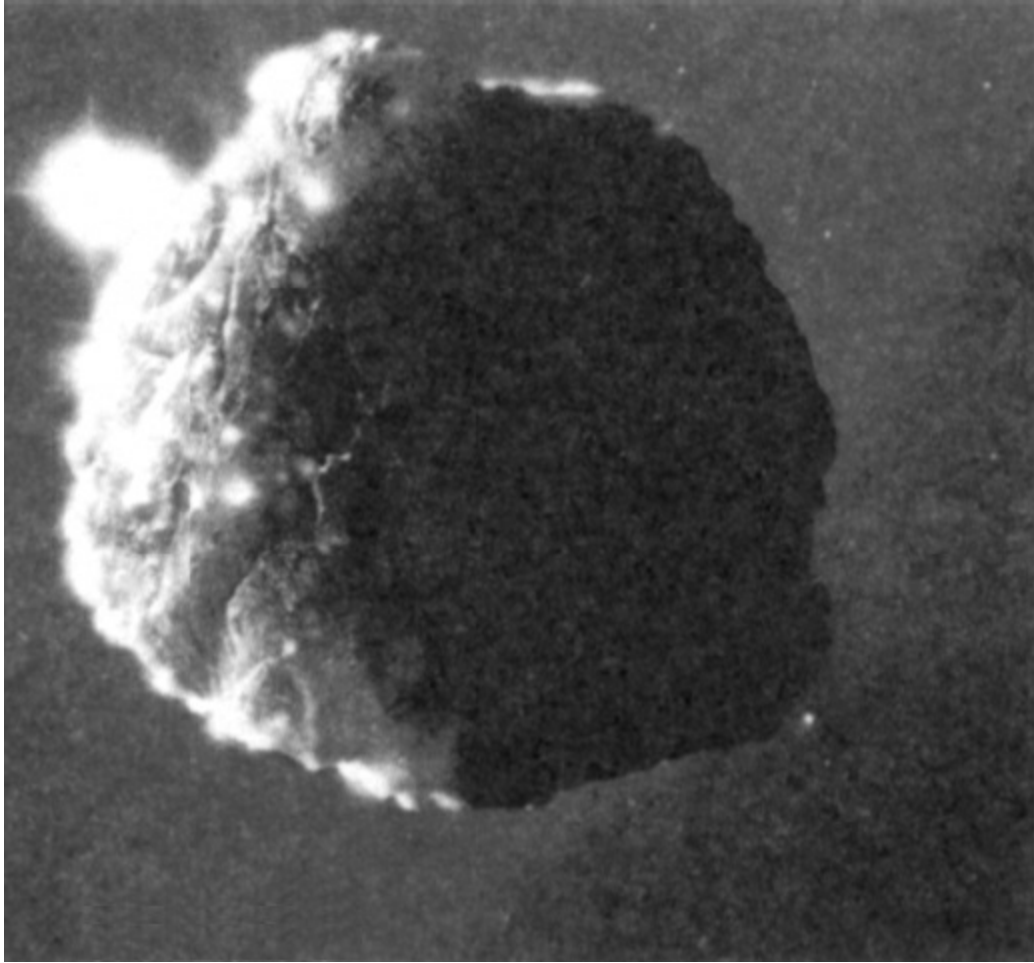


图9-30 彗星是一个脏雪球

在海王星的轨道之外，有一片彗星的仓库，存在很多的彗星，因为某种扰动，彗星就会从外边掉进来，穿过其他行星的轨道，朝太阳飞过去，然后再绕太阳一圈返回去。彗星的轨道有的是椭圆，有的是双曲线，有的是抛物线。轨道为双曲线和抛物线的彗星回去肯定就再也不会回来了，但是椭圆轨道的彗星还要回来，比如说哈雷彗星就会返回来。

当彗星飞近太阳的时候，被太阳上的光给照化了，照化了以后，气体就出来了。气体出来后，最初以为是光压，光子的压力，让气体背对着太阳出现一个尾巴。现在认为光压不够大，彗尾主要是太阳风造成的，太阳风就是太阳喷射的粒子流。太阳不断向外喷射粒子流，把彗星周围的气体向反方向压，所以彗尾总是背对着太阳。

彗星的撞击

彗星的尾巴很长。1910年的时候，大概是哈雷彗星，它的尾巴要扫过地球。当时有个天文学家知道了，饭都吃不下去了。觉得这下完了，这么亮的一个尾巴扫过去，不是会把我们给烧完了吗！结果没事儿。为什么呢？因为彗尾非常稀薄，比我们实验室抽的最好的真空还要空，所以扫过去没事儿。但是彗头撞上来可很厉害，我原来不太相信，以为彗头不过是一个冰球，它能怎么样。

1994年让地球人大开眼界，有一个彗星的头碎了，碎成二十多块砸在了木星上，在木星上面砸出一洞（图9-31）。砸的能量有多大？这二十几个碎块砸上去以后，相当于二十亿颗原子弹（以广岛那颗原子弹为准）那么大的威力。砸后的黑斑存在了好长时间。注意，木星是一个流体星，它不是固体星，但是砸完之后还是留下一个洞。在流体上砸的那个窟窿，可以把地球搁进去。所以彗头要是撞在地球上，事情还真是不可低估。

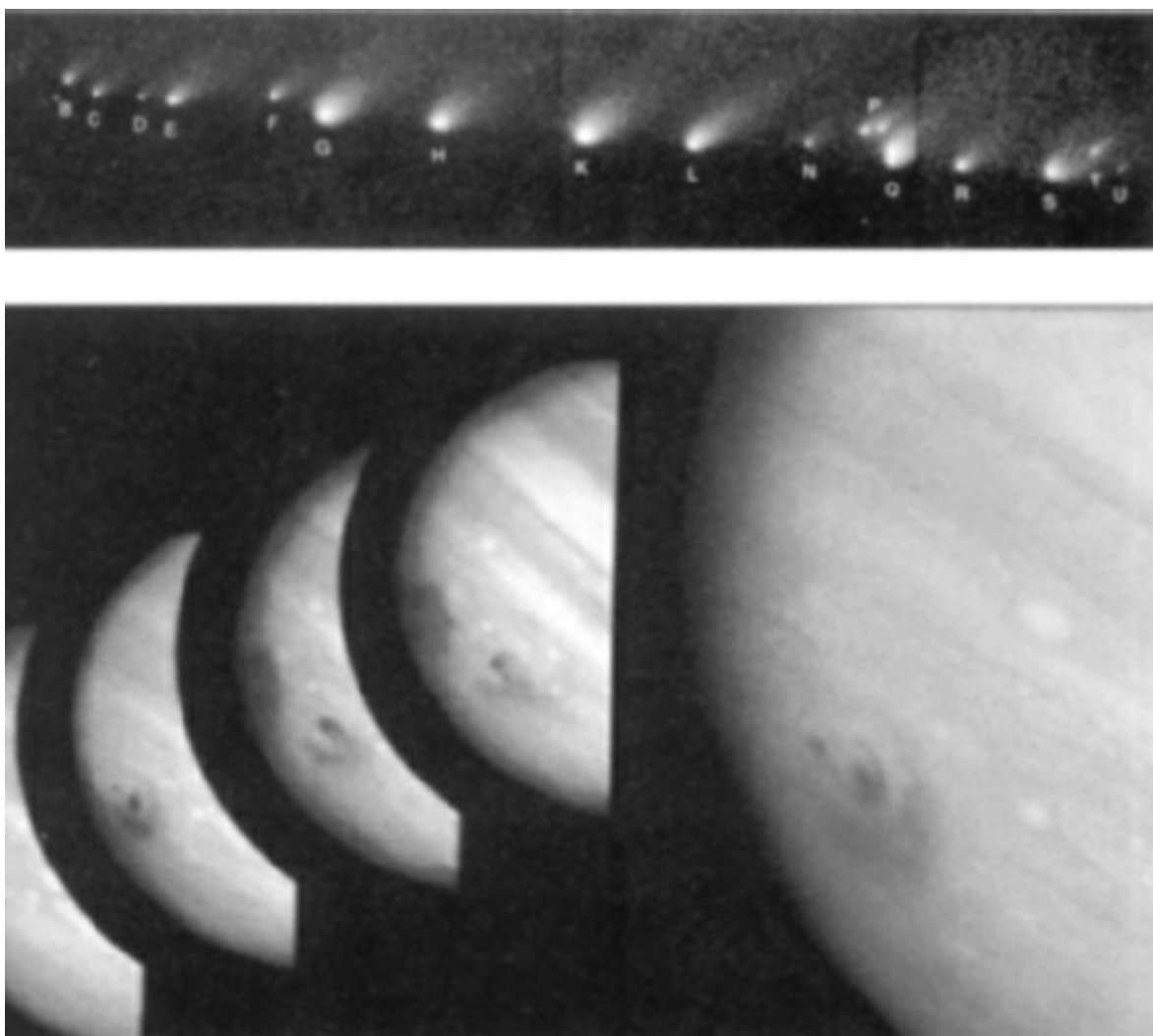


图9-31 彗木相撞

历史上的彗星

在人类历史上，彗星的出现是很引人注目的事情，往往都有一些记录，而且一般的民族都认为，彗星的出现是不吉利的，或者是要打仗了，或者是要有瘟疫了，反正是不吉利。中国人说彗星是扫帚星。

比如说汉朝时，有一本书叫《淮南子》，里面说，武王克商的时候，周武王的军队向商的首都殷和朝歌那一带前进的时候，出现了彗星。这彗星的“把”朝着殷人，殷人要倒霉了。

后来这段记录被用来研究武王克商的年份。因为中国最早的、比较完整的编年史是《资治通鉴》，《资治通鉴》是从公元前403年，韩、

赵、卫三分晋国开始往后记的。这一年也被现代历史学家认为是中国进入封建社会的开始。那以后的编年史非常清楚，一年一年非常清楚，以前的不是很清楚。当然还有其他很多史料可以追溯到公元前841年，但是更早以前的记录就不太清楚了。

近来开展的夏商周断代工程就考察这个问题。考察的时候，专家们注意到《淮南子》对彗星的这一记录。研究发现，如果这颗彗星是哈雷彗星的话，它出现的时间应该是公元前1057年。现在根据很多资料研究后，认为武王克商最可能的年份是公元前1046年，差一点。《淮南子》这本书，不是一本专门的史书，它是一本杂书，很有可能是在武王克商的前后有彗星出现。因为彗星出现是一件大事。当时来讲，不论老百姓，还是帝王，都很重视，都在考虑这东西是吉利还是不吉利，都在那儿猜。所以古人印象深刻。

天文用于考古

现在我们判定历史上很多事件的时候，往往要利用天文资料，利用天文学的一些记录和研究成果。

比如说，在判定中国古代历史的时候，有本叫《竹书纪年》的书很重要。《竹书纪年》很奇怪，对它的真伪长期存在争议。现在我们知道是魏国的史书，作为陪葬品埋在魏王墓里头，躲过了秦始皇的焚书坑儒。

后来在晋朝的时候，有一个盗墓贼，名叫不准。他把墓掘开去偷东西的时候，火把烧完了，没有照明的东西了，发现旁边有好多竹片，他就拿来点火照明。他把好东西拿走，走出来就把竹片一扔，逃跑了。当地的一些文人一看，这竹片上面有很多古字啊。因为是在秦始皇改革之前的那些文字，也不太认识。后来他们就把这个墓掘开了，用板车拉出好多车竹简。那些竹简上有许多历史记载。

比如说就记载了：周懿王元年，“天再旦于郑”。就是说周懿王的元年，在郑这个地方，也就是河南，天亮了两次。很多人都认为这是瞎

说，天怎么会亮两次呢。现代天文学家研究以后，认为如果是在太阳即将出来的时候，发生日全食，就会出现“天再旦”现象。然后中国人就研究了，研究这次日全食发生在什么时间，提出了一个年份。结果日本人提出不同意见：你们说得不对，你们说的那一次日全食，中国看不见，在太平洋上的人才能看见。

日本人认为中国能看到的那次日全食，应该发生在公元前899年。于是天文考古就确定了，周懿王元年是公元前899年，这就把中国历史上有确切纪年的时间推到了周懿王元年。

现在的确切纪年已上推到了武王克商的年份，即公元前1046年。《竹书纪年》上还有一段话很有用，说商把首都从别的地方迁到殷以后，273年没有动位置。由此可以把盘庚迁殷的年份定出来，在周武王克商以前273年。所以天文对历史研究很重要。

俄罗斯人也利用天文进行历史研究。历史上，蒙古军西征的时候，曾经跟俄罗斯联军进行了一场决战。俄罗斯联军全军覆没，这场败仗在什么时候打的不清楚。后来根据历史的记载，俄罗斯联军在渡过一条大河迎击蒙古军的时候出现了日全食，根据这次日全食就推算出了这一仗是哪年，甚至哪天打的。这在他们历史上是很重要的一件事。

5. 陨星与流星

通古斯陨星之谜

再给大家讲一颗陨星——通古斯陨星。通古斯这个地方位于西伯利亚，我看有的书上说通古斯就是东胡，但有的书上说不是，存在争议。不过西伯利亚就是鲜卑利亚，这一点比较确定，也就是西伯利亚原意为“鲜卑人的故乡”。

十月革命之前，在1908年7月的一天早晨，大概六七点钟的时候，在西伯利亚的上空出现了一颗火流星，轰轰响着从天空飞过（图9-32）。周围的一些居民（那里居民不是很多，不过还是散居了不少），

看到一个大火球从高空飞过去，进入到森林里，一声巨响，在那里升起了冲天的云柱。在西伯利亚大铁路火车上的人也都看见了。这个火流星咣的一声，砸在森林里。爆炸的规模非常之大。沙皇政府当时没有功夫管这个事儿。



图9-32 通古斯陨星

十月革命以后，苏联政府想起要研究这件事，就派了考察队到那儿去考察。找到了那个地方，爆炸的半径十几公里。在爆炸半径的范围内，所有的树木都被摧毁。图9-33就是撞击后的景象。



图9-33 撞击后的景象

大家对这颗陨星究竟是什么东西搞不清楚。有人认为这是一颗小行星撞过来了。但是在那里挖了很久也没有挖到什么。而且刚开始负责挖的这位科学家，因为卫国战争爆发参军去了，后来在前线牺牲了。战后又有了新的考察队来考察，在那个地方工作了很久，始终找不到陨星碎块。它要是个大陨石，砸了之后总会有东西啊！可是陨石坑底下和周围什么都没有留下，几乎什么东西都没有。树全都倒了，燃烧得很厉害。当时欧洲的很多地方，包括德国的气压计都测到了大气压强的变化，而且爆炸的空气震荡波绕地球一周后（30小时后），再次被德国波茨坦的气压计记录到，可见这次冲击是非常厉害的。

但为什么没有残留物呢？于是各种科幻小说就出来了。有人说，这是一艘外星人的原子飞船，在这儿失事了，外星人本来想要软着陆，没

想到硬着陆了，一下炸了。因为是核爆炸，所以后来没有什么东西留下来，这个地方也确实有点放射性，但放射性很弱很弱。黑洞理论出来以后，又有人说是小黑洞砸在那里面了，从这儿砸进去，从太平洋飞出去了。所以大家谁也没看见它出来。说什么的都有。现在大家比较相信的说法，认为可能是彗星，是彗星的脑袋。

一九六几年，苏联的天文学家就说这是彗星头。因为彗星的头一撞，它都是尘埃啊，冰啊什么的，当然就没留下渣子了。后来还有争议，现在大概争议比较小了。多数人认为这可能是彗星的撞击。

流星和流星雨

除去彗星以外，还有流星和流星雨。图9-34显示天空中有一颗流星划过。图9-35是流星雨。现在的流星雨都不是很壮观，说是流星雨，你等在那儿看，等半天才有一颗流星过来，再等半天，又有一颗。反正你得看半天，才能看到一颗。图9-35是一张流星雨的照片。图中这些弧线都不是流星，而是天上的恒星。这些划出直线的才是流星。拍摄的方法是把照相机固定，然后长时间曝光。由于天上的恒星都在围着北天极转，所以曝光以后，天上的恒星都画出弧线。在长时间曝光中，突然过来一颗很亮的流星瞬间划过天空，划出一段直线，相机拍了下来。再过来一个，又拍下来。



图9-34 流星



图9-35 狮子座流星雨

古代的时候，出现过密集的流星雨。图9-36是1833年的狮子座流星雨，这是一张版画。我相信当时规模可能是比较大，要不然不会叫它流星雨。如果等了五分钟才出现一颗，那叫雨吗？我有一次就打算看流星雨，等了好半天才看见一颗，后来也就不想看了。

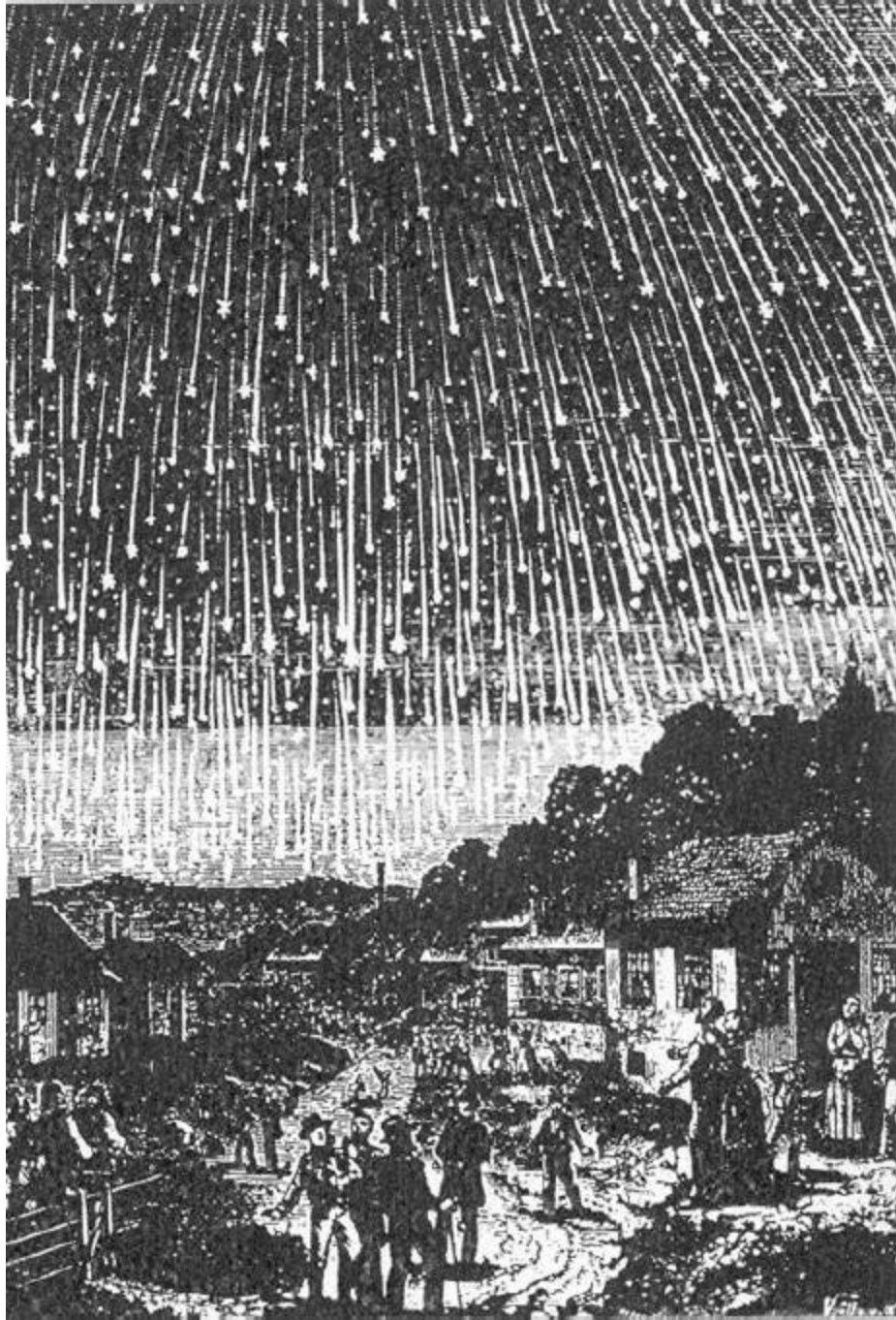


图9-36 1833年的狮子座流星雨

流星雨是什么，就是地球穿过流星群碰到的微粒。怎么会有流星群呢？流星群就是彗星头碎了以后形成的大量小碎片（图9-37）。形成流星雨的微粒在大气中都烧光了，一般不会落到地面上。大的彗头则不一样，那落下来是很厉害的。

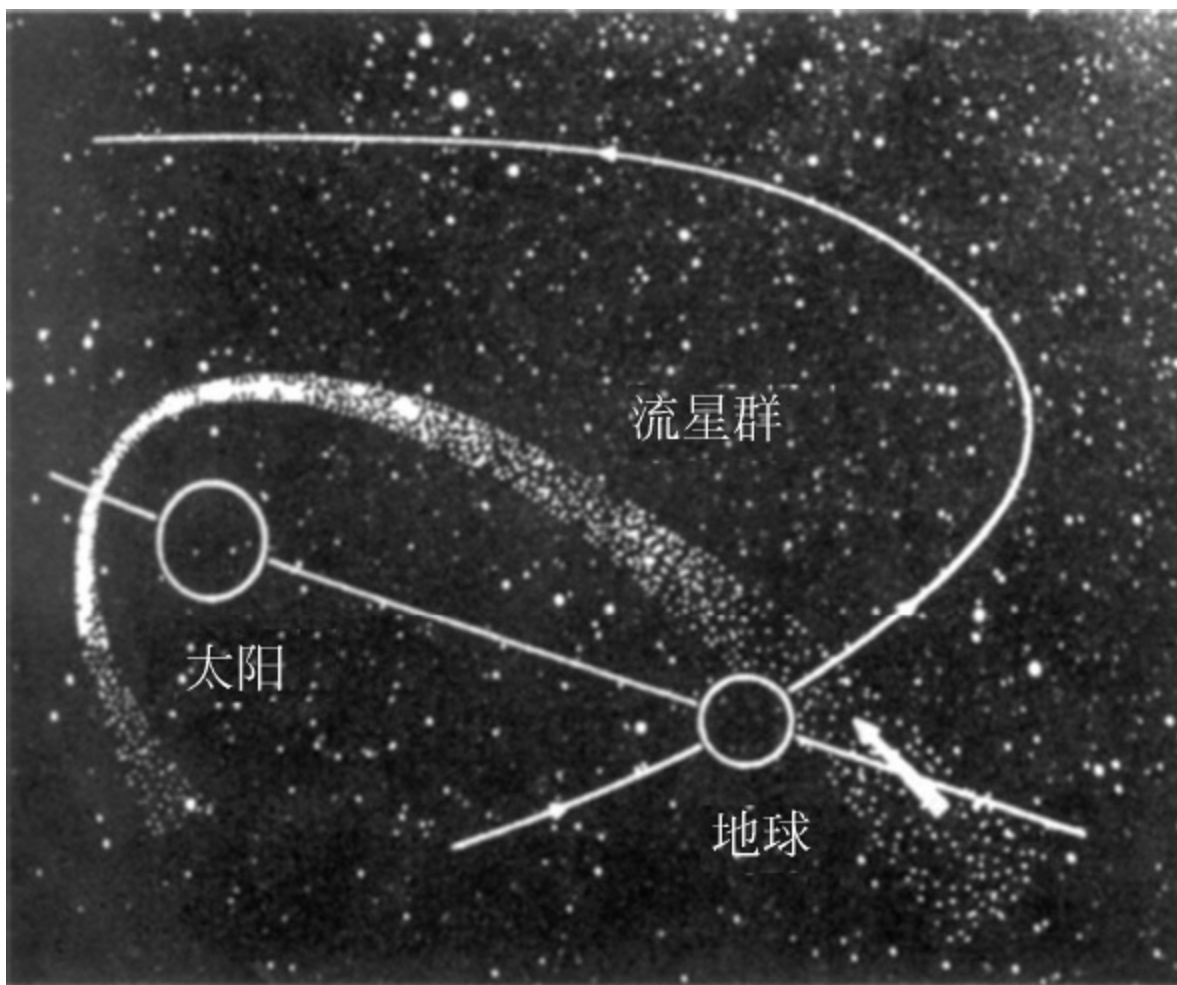


图9-37 流星群与地球轨道相交示意图

陨铁与陨石

图9-38是美国亚利桑那州的陨石坑。这块陨石好像是一块陨铁。美国的一些科学家感兴趣，想研究这个，他们觉得陨铁可能在坑下边，但又没钱去挖，就去跟钢铁大王商量，说：你们要是把这块铁全部挖出来，就不用再炼生铁了，这是多大的一块铁啊！



图9-38 亚利桑那州的陨石坑

企业老板就派了勘探人员来，钻了半天，没有！没有，他们也没有办法。后来科学家们又跟人家说，我们搞错了，不是在正下方，陨星可能是斜着撞过来的，陨铁可能位于坑的斜下方，你们再来吧，再来继续挖。

勘探人员又按照学者们的指点往下钻，钻了400多米深后，钻头似乎碰到了铁。但是这么深的铁如何才能取出来呢？商业价值已经没有了，资本家们失去了兴趣，不再出资了，因此这块可能存在的陨铁至今还睡在地下深处。

天上掉下来的陨星主要是铁。石质的陨石比较少。图9-38是亚利桑那州的陨石坑，你看在太阳照耀下多美啊。所以你们如果到美国去的话，可以到那里看看。图9-39是新疆的一块陨铁，重32吨。图9-40是南

非的霍巴陨铁，重60吨，是已经发现的世界最大的陨铁。图9-41是陨石。



图9-39 新疆陨铁



图9-40 南非霍巴陨铁



图9-41 陨石

6. 天文观测简介

太阳系有多大！

我再简单回顾一下太阳系，地球的半径是6400公里，月球的半径是1700公里，太阳的半径是70万公里，日地距离是1.5亿公里，叫一个天文单位。光大概走七八分钟的样子，就可以从太阳到达地球。冥王星距太阳是40个天文单位，光走5小时。太阳系的直径是多少？大概是1光

年。

太阳的引力的控制范围（半径）大概是半光年。所以我们现在还不会有飞行器飞出太阳系，最多也就是飞过冥王星的轨道。要想真正飞出太阳系，还需要做更多的努力。

天文台与望远镜

简单说几句天文仪器。图9-42是天文望远镜，咱们北师大也有天文望远镜。科技楼上有一个，物理楼上也有一个。上面的圆顶是可以向两边移动张开的，张开后，望远镜就露出来了。我给大家看的资料不是最新的。我想将来会有很多搞天文的同志来做讲座，你们会从他们那里看到许多新资料、新图片。图9-43是接收无线电波的装置，叫射电望远镜。它组成一个阵，有时是成十字的阵，组成阵列。还有一种是利用地形修的，反射型的接收无线电的天文望远镜。图9-44是波多黎哥的直径308米的巨型射电望远镜。它就是利用山谷地形修建的。中国正在贵州利用山谷地形修一个直径500米的世界最大的巨型射电望远镜，因为那边的山谷地形和观测条件也不错。

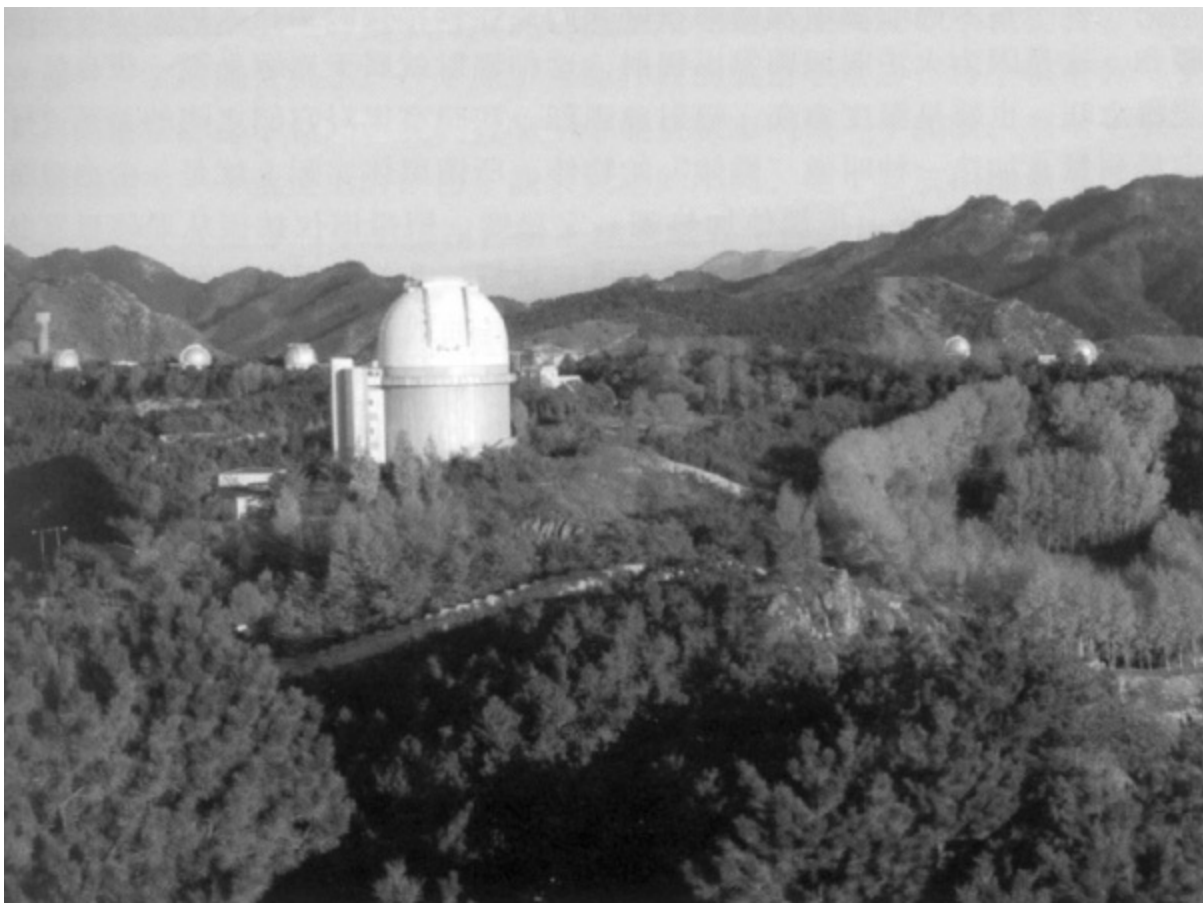


图9-42 国家天文台的2.16米望远镜

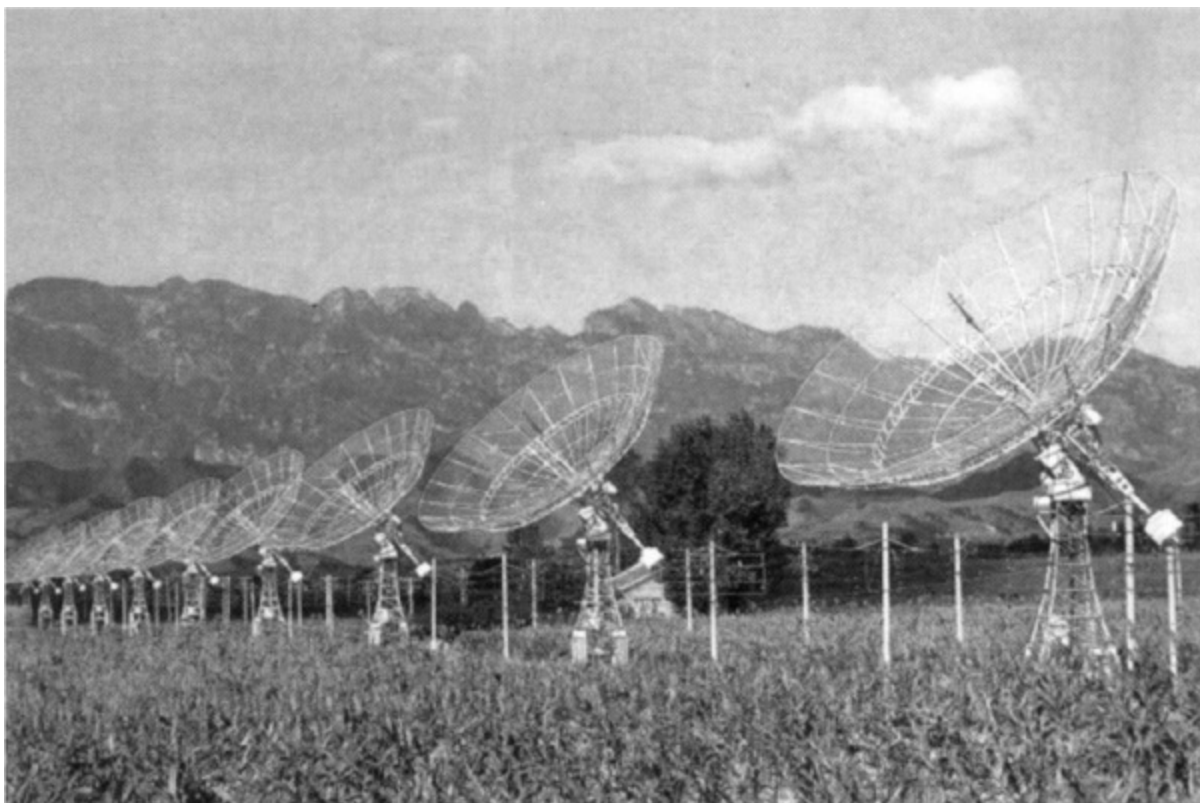


图9-43 密云射电干涉仪

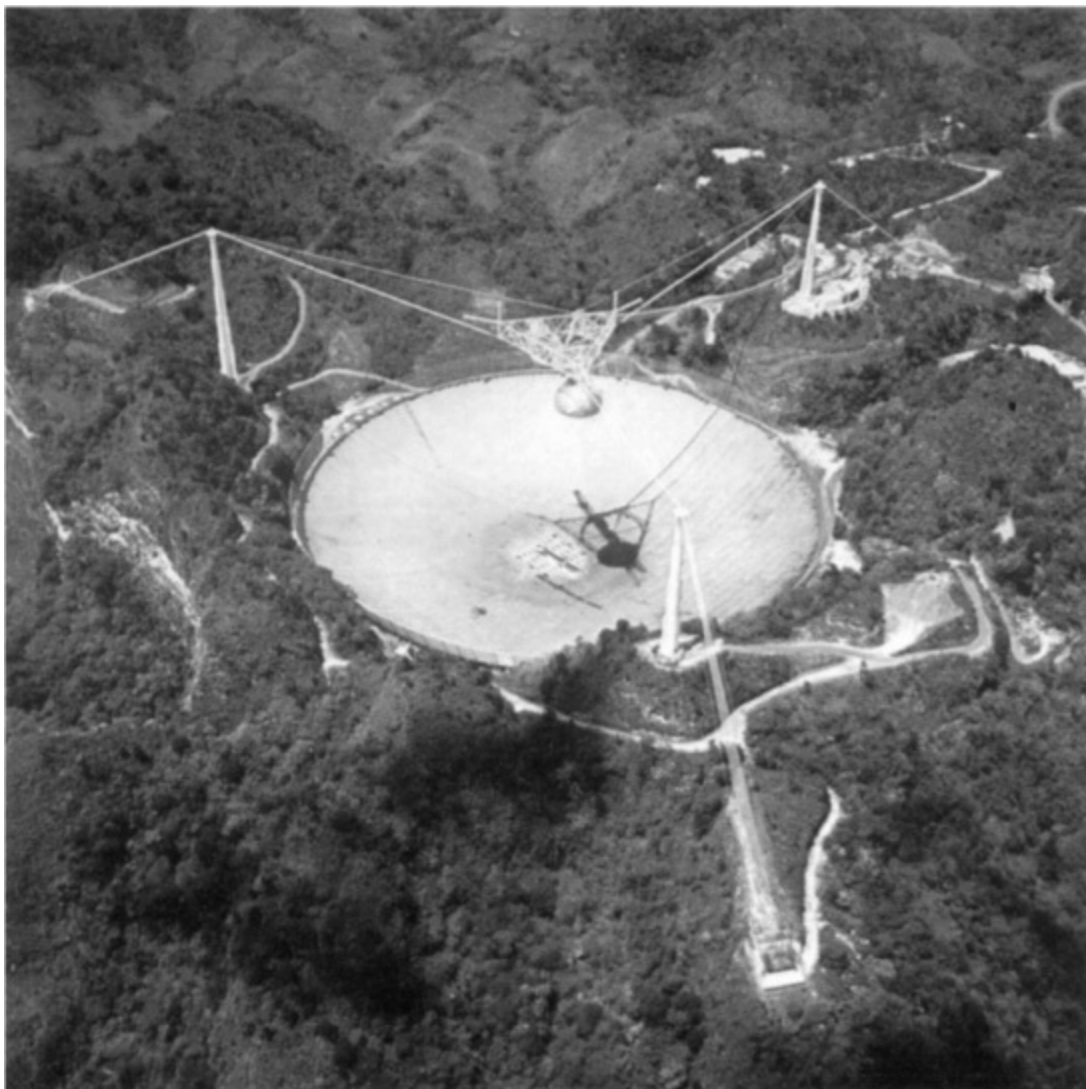


图9-44 波多黎哥的巨型射电望远镜

我还要重复一句，望远镜不止是在看远方，而且是在看历史。你看到的太阳是8分钟以前的太阳。太阳要是炸了，我们8分钟以后才知道它炸了。天狼星距离我们9光年，是离我们比较近的一颗恒星，它的光到达地球要走9年，所以我们看到的天狼星是9年前的模样，如果它出现问题，我们9年以后才会知道。

我今天就讲到这儿，有关的资料在《物理学与人类文明十六讲》里有，高教社出版的这本书。你们可以看到各种天文书籍，有问题也可以问问天文系的老师，他们会给你们介绍各种各样的天文读物。现在讲天

文的书和杂志很多。一般人对天文都比较感兴趣。

第十讲 时间之谜

现在我们来讲《时间之谜》。先请大家看一首词：

忆昔午桥桥上饮，
坐中多是豪英。
长沟流月去无声，
杏花疏影里，
吹笛到天明。

这首词怎么样？美不美？（听众齐声回答：美！）我青年时代看到这首词就非常欣赏，所以我在自己写的第一本书《探求上帝的秘密》里，就把这段诗搁在前面了。其中一个用意是告诉读者，我这本书上讲的可都是豪英啊！另外一个用意是让大家看，“长沟流月去无声”，非常像时间在流逝。我引用这首诗有双重含义。

作者陈与义是宋朝人。南宋还是北宋呢？生于北宋，死于南宋。这个人在宋朝的时候当过吏部侍郎，后来当到参知政事。吏部侍郎就是组织部副部长。侍郎就是副部长，尚书是部长，参知政事就是副宰相、副总理这个级别的官。陈与义官当得还是不小的。

1. 时间是什么？

这一讲我主要讲时间。大家看，第一个问题：“时间是什么？”这个问题不好回答。第二个问题：“时间有没有开始与终结？”这是因为彭罗斯和霍金在其中做了研究工作，所以我从一个学物理的人、学相对论的人的角度，来给大家介绍他们究竟证明了什么，以及这个问题究竟应该怎么解决。另外一个问题就是：“时间怎么测量？”你们会看到，时间测

量也绝非易事，下面我就给大家讲。你们会觉得这些怎么都成了问题！看来不像问题，但是又确实都是问题。

你们看，一位生活在中世纪的基督教著名学者圣·奥古斯丁说：

时间是什么？人不问我，我很清楚；一旦问起，我便茫然。

你们想是不是这样？这个问题真是不好回答。

百花齐放，百家争鸣

首先给大家介绍一下，人类文明出现以后，在出现思想家之后，就有人开始思考时间问题了，比如柏拉图等人。

在人类文明史上，根据现在的记载，什么时候就出现了非常杰出的思想家呢？研究表明，从公元前六百年到公元前三百多年这一段时间，是人类历史上非常重要的时期。

公元前六百年，诞生了犹太教。当时犹太人被抓到了巴比伦城下，逼着他们住在那儿，成为“巴比伦之囚”。当时的犹太人亡了国，他们不得不生活在那个地方，但他们相信耶和华会派人来救他们，也就是他们心目中的上帝会来救他们。于是在那个地方诞生了犹太教，那是公元前六百年左右的事。犹太教是天主教、基督教和东正教的鼻祖。后边的这些宗教都来自于犹太教。伊斯兰教也受它们的影响。

公元前五百多年，对于中国来说是非常重要的时期。历史上影响中国人思想的三个最伟大的思想家，都生活在那个时代。一个是孔夫子，一个是老子，还有一个是释迦牟尼。对我们中国人影响最大的宗教是佛教和道教。佛家又叫释家。道教呢，它比较牵强地把自己跟老子联系起来。老子确实是老庄学说的创始人，道家学说是他最早提出来的。孔夫子的思想不是宗教，但是有点像宗教，有人叫它儒教，其实就是孔夫子的儒家。这是对中国影响最大的三套学术理论。

这一历史时期，东、西方都处在战国时代，伴随着刀光剑影，都出

现了百花齐放，百家争鸣的局面。

在我们中国，诸子百家出现在春秋到战国的那段时间，公元前五百多年到公元前两三百年的时候。

同样的，在西方，希腊也出现了很多的思想家。比如说一个重要的思想家苏格拉底。苏格拉底是一位雄辩家，主张跟学生自由讨论问题。他遭到政敌的攻击，说他是“无神论者”。“无神论者”这个罪状当时很大啊。同时说他败坏青年，最后把他处以死刑，逼他服毒酒自杀。他服毒酒自杀的时候，看守他的人都想让他跑。说你要逃走，我们不管。但是他还是自杀了，他没有跑。

苏格拉底有一个非常杰出的学生，叫柏拉图。据说，他在收柏拉图做学生的头一天晚上，梦见有一只小天鹅落在了他的膝盖上，很快地羽毛就丰满起来，然后唱着优美动听的歌飞上了蓝天。第二天就见到了柏拉图，收了柏拉图这个学生。苏格拉底的著作不是他本人写的，是他的学生整理的。这点很像我们孔夫子的《论语》，不是本人写的，而是后来他的学生根据记录整理出来的。

柏拉图是非常著名的思想家。大家经常谈到的，就是他曾经提到过的理想国，也就是乌托邦。那当然是奴隶社会的理想国。他认为理想的国家应该由哲学家来治理，由哲学家来制定法律，除去哲学家以外，所有的人都必须遵守法律。他认为这是最理想的国家。

柏拉图还提到大西洲的事情。说在海峡的对面有一个繁荣的国家，后来由于火山爆发和地震，一下子就淹没在大海的波涛之中了。这个海峡的对面，指的是哪个海峡？没有说清楚，这个故事对西方影响很大。我们中国人现在也听说了这个大西洲的故事。

据说柏拉图本来很关心政治，后来在他的老师苏格拉底被诬陷死亡以后，他对政治大失所望，于是就远离政治。他周游列国到了埃及，在埃及的神庙里，听祭司给他讲了有关大西洲的故事，他把它记录下来。

柏拉图：时间是永恒的映像

现在来讲柏拉图的哲学观点。柏拉图对时间有一个论述，说：时间是“永恒”的“映像”。这是怎么回事呢？柏拉图认为：真实的“实在世界”是“理念”，是一种叫理念的抽象东西。我们接触到的万物和整个宇宙，都不是真实、实在的东西，都不过是“理念”的“影子”。“理念”完美而永恒，它不存在于宇宙和时空中。而万物和宇宙是不完美的，是在不断变化的。

大家注意，“理念”是真实存在的东西。我们看见的万物都不过是“理念”的影子。“理念”是永恒的，而万物是变化的。那么理念在永恒，万物在变化，万物就需要有一个跟永恒对应的东西，那就是时间。

所以“时间是永恒的映像”。看看，很深奥啊，是不是！我看了好多次，都看不懂怎么回事儿。造物主给“永恒”创造了一个“动态相似物”，就是时间。

总的来说，柏拉图的观点就是：时间是“永恒”的映像；时间是“永恒”的动态相似物；时间不停地流逝着，模仿着“永恒”；时间无始无终，循环流逝，大概三万六千年是一个周期。希腊人的时间观是周期的，柏拉图和他的很多优秀的学生都这样认为。

亚里士多德：时间是运动的计数

亚里士多德是柏拉图最优秀的学生，但不是他的学说最满意的继承人。因为亚里士多德把柏拉图的学说颠倒了一个个儿。他认为我们看不见的东西都不是真实的，我们看见的万物才是真实的东西，他不承认有“理念”这个东西。从我们今天的观点来看，柏拉图的理论是唯心的，亚里士多德的理论是唯物的。

亚里士多德这个人学问非常大。柏拉图曾经创立了一个叫阿卡德米的学院，英文的academy这个单词就是从这儿来的，学院这个词就是从这儿来的。而亚里士多德又建立了一个吕克昂学院，也就是所谓逍遥学院。亚里士多德愿意一边走，一边跟他的学生讨论，因此历史上管他们

叫逍遥学派。亚里士多德这个人非常聪明，个子不高，说话尖刻，经常得罪人。他对时间有一个论述，说：时间是运动的计数；时间是运动持续的量度；他也认为时间是循环的。这点跟他的老师是一致的。这就是亚里士多德的时间观。

走向统一的世界

我还要对亚里士多德再介绍两句。亚里士多德是西方著名的君主亚历山大大帝的老师。亚历山大大帝的父亲——腓力二世，是马其顿的国王，腓力请亚里士多德给自己的几个孩子当老师，包括亚历山大。不久之后，腓力二世被刺死。据说是亚历山大的母亲派去的刺客，那是因为他又娶了一位年轻漂亮的王妃，威胁到亚历山大母子的利益与安全。腓力二世被刺以后，亚历山大就继承了王位，做了马其顿的国王。

亚历山大很能干，很快就统一了希腊。然后进入小亚细亚地区，进入阿拉伯半岛，进入了亚洲。后来又进入非洲。军队还一直往东前进到达印度河流域。他在公元前三百多年时建立了一个横跨欧、亚、非三洲的大帝国。

当亚历山大走到埃及北部海岸的时候，他站在地中海边，遥望对面的希腊故乡，下令在这个地方建一座城市，以他的名字命名；又下令在这儿修建一座灯塔，也以他的名字命名。让灯塔遥对着大海对面的希腊，他要让航船在老远的地方就能看到这座灯塔，这就是著名的世界七大奇迹之一的亚历山大大灯塔。

亚历山大死后，他的部将托勒密在埃及建立了一个希腊人为统治民族，埃及人为被统治民族的国家，并按照大帝的遗愿，修建了亚历山大城，修建了亚历山大大灯塔。而且托勒密一世、托勒密二世热爱科学，他们还建立了亚历山大科学院和图书馆。亚历山大大帝的遗体就被安葬在那座城市。

亚历山大统一西方是在公元前三百多年。过了不到一百年，阿育王统一了印度，他推崇佛教。阿育王死后十一年，秦始皇统一了中国。这

三块人类文明最早发展的地区，基本都在公元前二三百年的时候分别实现了统一。

孔子：逝者如斯夫

再来看中国古代的时间观。我不是专门研究历史和哲学的，只能就我所知道的东西讲一讲。《论语》上面讲：

子在川上曰：逝者如斯夫！不舍昼夜。

什么意思呢？就是说孔夫子坐在河边上讲，时间，就像这河流一样永远不停地流逝。孔夫子把时间叫做“逝者”，这是很高明的！因为时间除去有一个可以测量的特性以外，它还有一个流逝的特性。时间是不可逆的。对不对？这句话里包含流逝、发展，有这层含义。所以他这样讲是很高明的。

螺旋发展的时间

中国古代受三种思想的影响，一个是释家（佛教），一个是道家，一个是孔夫子即儒家，所以中国古代知识分子的思想是比较复杂的。一般来说，他们认为时间是有周期的，但不是简单的重复，而是螺旋发展的。比如唐朝刘庭芝的诗：

年年岁岁花相似，
岁岁年年人不同。

从中可以看到时间周期的相似性和不断发展性。更妙的是宋朝晏殊的词：

一曲新词酒一杯，
去年天气旧亭台，

夕阳西下几时回。
无可奈何花落去，
似曾相识燕归来，
小园香径独徘徊。

这首词相当美，是吧！“去年天气旧亭台”，这是讲循环的相似；“夕阳西下几时回”，时间不停地向前流逝；“无可奈何花落去”，万物都必须与时俱进；“似曾相识燕归来”，循环的相似。这里最美的两句就是：“无可奈何花落去，似曾相识燕归来。”

牛顿的时间观：永远平静流逝的河流

我们再来看看近代西方的时间观。大家都知道，牛顿认为有一个绝对空间和一个绝对时间，不过他还认为存在相对空间和相对时间，但他认为绝对空间和绝对时间是根本的。绝对空间就像一个空的箱子，绝对时间就像河流一样流逝。时间和空间是各自独立的，时空和物质及其运动也没有联系，这就是牛顿头脑中的时间与空间。牛顿曾经用水桶实验论证过绝对空间的存在，但是没有论证过绝对时间的存在。任何人都没有论证过绝对时间的存在。

按照牛顿的观点，时间是均匀的，有方向的，没有起点和终点的，是永远存在的“河流”，没有涨落也没有波涛。如果物质消失了，时间和空间还会继续存在。这是牛顿的想法。

不同的声音

和牛顿同时代的莱布尼茨说：时间和空间都是相对的；空间是物体和现象有序性的一种表现；时间是相继发生的事件的罗列；没有物质的话，根本就没有时间和空间；不存在脱离物理实体的时间和空间。这种观点跟牛顿完全对立。这两种观点一直持续到现在都存在。

到了爱因斯坦那个时代，马赫对牛顿的观点提出了尖锐的批判，他认为牛顿的说法不对。马赫认为不存在绝对时间和绝对空间。马赫这个

人，就物理水平来说，只能算一个三流物理学家，不是特别棒的，当然也有成就。但是他敢说祖师爷不对，这也很厉害。他的观念对爱因斯坦影响极大，爱因斯坦认为，正是马赫对牛顿绝对时空观的批判，引导自己创建了狭义相对论和广义相对论。

时空的维数

时间是几维的呢？跟牛顿同时代的哲学家——洛克说：时间是一维的。时间为什么是一维的呢？不清楚。反正大家觉得时间是一维的，你也想象不出它是二维或更高维的情况。

空间是几维的呢？空间是三维的，为什么呢？这倒是有一个论证。库仑定律表明，力和距离的平方成反比。如果力和距离的平方成反比的话，空间就一定是三维的，这是李政道先生讲过的。所以，对于“空间是三维的”是有实验支持的，那就是库仑定律，力和距离的平方成反比。

爱因斯坦：时空是一个可弯曲的整体

爱因斯坦的狭义相对论认为，时间和空间是一个整体，是不能分割的。能量和动量也是一个整体，也是不能分割的。他开始把时空连在一起了，叫四维时空；把能量动量连在一起了，叫四维动量。在相对论的研究中，常用这两个概念。

爱因斯坦的广义相对论进一步认为：时空和能量动量之间存在关系。能量动量的存在，造成了时空的弯曲，时空的弯曲反过来影响物质的运动。但是在爱因斯坦的理论中，如果物质消失了，时空依然存在，只不过时空从弯曲的变成了平直的。这是爱因斯坦相对论最后的结果。

爱因斯坦：时空是物质广延性的表现

不过，爱因斯坦晚年的时候对自己的这种描述产生了怀疑。爱因斯坦本人没有写过几本书，他写过一本科普读物，叫《狭义与广义相对论浅说》，还写过一本学术著作叫《相对论的意义》，大概一共就两三本。他主要是发表论文。

爱因斯坦在《狭义与广义相对论浅说》1952年第15版的说明里说了一句话：“空间和时间未必能看作是，可以脱离物质世界的真实客体而独立存在的东西。并不是物体存在于空间中，而是这些物体具有空间广延性。这样看来，‘关于一无所有的空间’，就失去了意义。”大家注意，他的物理理论（相对论）认为物质可以跟时空脱离。物质没有了，时空依然存在，只不过时空变平了，不再弯曲了。现在，他怀疑自己的这个理论，觉得自己的理论还应该进一步往前发展。

在1955年逝世前，爱因斯坦一直坚持认为，时空是物质伸张性和广延性的表现；不存在一无所有的时空；时空应该和物质同生同灭。这个观点跟莱布尼茨的观点有些接近了。他最初的想法体现在狭义相对论与广义相对论中，跟牛顿的观点接近，然后新的哲学思考又跟莱布尼茨的观点接近。我们看到，伟大学者的思想不是永恒不变的，而是在思考中不断发展的。伟大思想家的观点对后世的影响可以延伸很久很久。

量子引力：时间的泡沫与浪花

现在，物理学家正在试图把引力场量子化。大家知道，电磁场、电子场等物质场都已量子化了，而且都很成功。唯独引力场的量子化遇到了极大困难。引力场是时空弯曲的表现，它能不能量子化呢，一直有人在尝试做这方面的工作，但到现在为止都不成功。引力场量子化的各种方案都能取得一定进展，但是最后都会遇到一些难以克服的困难。

所谓量子引力，就是把引力场量子化。在这个理论中，时空是和物质同生同灭的。确实能做到和物质同生同灭，但是又有其他很多困难。比如说，时空在大范围内，我们看着是很平的，就像图6-14上面这个图。但是在很小的范围， 10^{-30}cm ，你就会觉得它有一点波浪式的起伏。 10^{-13}cm 是原子核的大小， 10^{-30}cm 就更小了。如果是 10^{-33}cm 那就非常非常小了。在这么小的范围内，时空泡沫和浪花之类的东西就都会显现出来（图6-14）。时空绝不像大家原来想的那么平滑。

2. 时间有没有开始与终结？

我们现在来介绍另外一个问题，就是时间有没有开始和终结。这个问题，自古以来就有一些聪明人讨论，但那都是些神学家和哲学家。

现在搞物理的人出来说了，说时间有开始和结束，所以这个问题就应该引起科学界的重视了。我对这个问题感兴趣也是因为彭罗斯和霍金这些搞物理的人出来研究这一问题。

时空的奇点

时间有没有开始和终结。这个问题是怎么在物理界引起注意的呢？是从研究时空曲率的奇点引起的。

相对论中早就发现，在静态球对称黑洞的中心 $r=0$ 的地方，有一个奇点，那个地方时空弯曲的曲率是无穷大，物质的密度也是无穷大，所以是一个奇点。在旋转的黑洞中心，有一个奇环（图4-2），那里密度是无穷大，时空的弯曲程度也是无穷大。

大爆炸的宇宙理论有一个初始奇点。如果这个宇宙将来会塌缩回来的话，还会有一个大终结的奇点。广义相对论中最重要的几个解都有奇点。我们知道，点电荷的密度也是无穷大，所以奇点问题并不是只有广义相对论才有的，其他的物理理论也有，只要模型太理想化了就会有。

那么奇点是不是广义相对论本身必然有的东西呢？对于广义相对论中的奇点，有两种看法。一种是苏联的几个物理学家认为，广义相对论中的奇点是因为我们把时空的对称性想得太好了所致，如果时空对称性不是太好，就不会形成奇点。比如说一个星体塌缩，如果不是标准的球对称塌缩的话，构成星体的物质就会在中间错过去，就不会集中形成奇点。如果星体不是标准的旋转对称地塌缩，那它也不会形成奇环。

奇点：时间的起点与终点

但是英国的数学家——彭罗斯，认为这种看法是不对的。他认为，奇点是广义相对论理论本身造成的，是不可避免的。他对这个问题进行

了深入研究，而且，他把奇点看成是时间开始和结束的地方，这是他很重要的贡献。

什么意思呢，比如说黑洞的中心有一个奇点，如果你把这个奇点挖掉，时空还是奇异的吗？这个时空是不是就没有奇异性了？他说：不是。挖掉奇点后，时空仍然有奇异性。因为你挖掉之后，时空会留下一个“洞”，这个窟窿你补不上。你补上去，奇点就恢复了。你要是在时空中把它挖掉的话，任何一根曲线到这儿就断掉了。所以他把描述时间发展的曲线看得很重要，时间发展的曲线会不会断掉，非常重要。他认为有奇点的话，时间发展的曲线就会断掉。这条曲线代表的时间过程就走到了尽头，所以他就提出了一个定理，并给出了一个证明。因为时间关系，我就不多讲证明过程了。

这个定理大体上是这样的：一个时空如果它的因果性是正常的，如果它有一点能量，如果它有一点物质，如果广义相对论是正确的，只要满足这些看来合理的条件，就一定会有一个过程，时间是有开始的，或者有结束的，或者是既有开始又有结束的。也就是说时空必定至少有一个奇点。

这个定理的证明是没有问题的。你如果仔细想，这个定理的结论还真是一个大问题，因为它等于证明了一个合理的物理时空，一定有一个时间有开始或结束的过程。这是物理学家对时间有没有开始和结束的第一次表态。这个问题当然很重要。

能否避免奇点？

但是为什么会出现这种结论呢？绝大多数人都认为，这是因为在做这个证明的时候，没有把引力场量子化。如果你把引力场量子化的话，就不会出现这个结论。引力场量子化就肯定能避免奇点出现。

我对这个问题很感兴趣，我注意到一个现象，就是：凡是有奇点出现的时候，都伴随着温度出现发散，或者是温度出现绝对零度。所以我非常怀疑，奇点定理的证明是违背了热力学第三定律的。我曾经在一些

论文中讨论了这个问题，我和我的研究生田贵花等人曾经发了一些论文，专门讨论这个东西。但是目前这个问题还没有解决，各个人往往持不同的观点。

我的观点是：如果热力学第三定律正确的话，时间应该没有开始和结束。就是说如果不能通过有限次操作，使系统温度达到绝对零度或升高到无穷大的话，时间就不应该有开始和结束。这是我的一个猜想，据此我也作了很多论证。这个证明是比较困难的，要用微分几何来求证。

当我们的第一篇论文投到美国的《物理评论》以后，那个杂志社说：你们这篇论文，更像数学论文。所以就让我们改投美国的《数学物理杂志》。改投之后，马上就登出来了。这个问题，我这次就谈这么多。

3. 时间的测量

时间的流逝性与测度性

现在来讨论第三个问题：时间测量的问题。时间有两个基本性质，一个是测度性，就是它能不能测量；再有一个就是它的流逝性。现在我们谈到时间性质的时候，往往着眼点都在它的流逝性上。流逝性是跟热力学第二定律有关的，就是表示时间总是不停地、有方向地、不可逆地向前流逝着。

我们知道，在一个孤立系统中，熵总是增加的，在一个绝热的过程中，熵也总是增加的；就是说时间发展有一个方向。大量搞物理的人都在讨论时间的流逝性。但是我现在更感兴趣的是时间的测度性。这方面我有点新的想法。

测度性是什么呢？就是怎么测量时间。我们都知道，用周期运动可以测量时间，这是自古就知道的。比如说，地球的周期运动，自转一圈就是一天，自转两圈就是两天，围太阳公转一圈就是一年。单摆，摆动一下就是一个时间单元，再摆动一下又是一个时间单元。

用周期运动测量时间带来的疑问

但是有人提了一个问题：你怎么知道周期运动中的第一个周期和第二个周期是严格相等的呢？长度的相等你容易弄清楚。比如这一段长度和那一段长度是否相等，可以拿一根尺子，量量这段，然后再量量那段，容易弄清这两段是否相等。

但是时间，你能把已经过去的周期挪回来吗？你能把未来的周期也挪过来吗？你没法做到这一点。所以相继的时间段是否相等的问题，是个不能证明的问题。跟牛顿同时代的哲学家洛克就认为，这是不能证明的，你只能规定它们相等。我们只能认为它们相等，实际上是不能证明的。

但是我们搞物理的人对这样的回答无法满足，还是希望知道对这个问题有没有什么办法解决。而且那个时候还产生了一些很混乱的思想，有些哲学家认为时间跟空间不一样，时间带有主观成分，它属于意识的范围，不属于物质。他们认为时间的周期是否相等只能凭直觉，两个钟是否对准了时刻，也只能凭直觉。在相对论诞生之前，这个问题就引起过一些争论。

异地时钟的校准：先约定光速

相对论诞生的前夜，数学家庞加莱曾经谈过：时间必须变成可测量的东西，不能被测量的东西不能成为科学的对象。庞加莱认为时间的测量有两个方面：一个是两个地方的钟，你怎么把它们对好，怎么校准，也就是怎么让它们同时或者同步；再有一个就是我前面说的，同一个钟的第一个周期和第二个周期，你怎么知道它们相等，你要想办法去证明这一点。

庞加莱当时认为，这两个问题还真不好办，他在自己的书中就讲：设想巴黎有一个钟，柏林也有一个钟，怎么把这两个钟对好。他说，我可以拍一个电报过去，那时还没有电话。拍个电报过去，告诉柏林的人，巴黎现在是几点，柏林的人赶紧把钟拧到那个时刻，这不就对好了

吗！这件事说来容易，可是你想过没有，这个电报从巴黎传到柏林还是需要时间的。你知道传过去用了多少时间吗？你要知道电报从巴黎传到柏林所需要的时间，就要首先把两个地方的钟对好，你对好两个钟又需要先知道电报传递的时间，所以这个问题构成逻辑循环，没法儿精确解决。

那么怎样才能解决呢？他说这恐怕要事先有一个“规定”（即“约定”）。我们知道，现在传播速度最快的是电磁波，也就是光波。我们可以规定光从巴黎传到柏林，和从柏林传到巴黎，所用的时间是相同的，可以做这个“规定”，规定真空中的光速是各向同性的。有人说：你怎么知道光速各向同性？我并不知道，但是我可以这样“规定”。许多学者仔细研究过这个问题，结论是：只能“规定”。我们规定了真空中光在两个方向的速度是一样的，然后就可以对钟了，就可以把钟校准。他谈了这个问题，但是我没有看到他的书中有特别详细的探讨。

庞加莱对爱因斯坦的启发

真正看到的详细讨论是在爱因斯坦的文章中。爱因斯坦应该知道庞加莱的上述想法。因为在相对论诞生之前，他和他的朋友在“奥林匹亚科学院”的活动中，曾读到过庞加莱的书《科学与假设》。很可能也读过庞加莱的另一篇文章《时间的测量》。

在这篇文章中，庞加莱写了自己对时间校准的上述想法。但是爱因斯坦没有谈他知道庞加莱的这一想法。这个情况可能比较复杂，因为他俩关系不是太好。爱因斯坦本来很指望庞加莱支持他的相对论，但庞加莱没有支持，而且对爱因斯坦评价不是很高。

有一次开会，他跟庞加莱见了面，他与庞加莱可能只见过那一次。回来后他就沮丧地对朋友们说：“庞加莱根本不懂相对论。”大概他跟庞加莱谈过，庞加莱没有支持相对论。

欧拉的“好钟”

还有一个问题，就是相继的时间段，怎么知道它们相等。另外一位

数学家欧拉，在庞加莱之前就想了一个办法，这个办法是比较怪异的。他说，怎么才能知道钟的第一个周期和第二个周期走得严格相等呢？可以这样判断：作为一个公理，我们可以认为，或者规定惯性定律是正确的，然后就用这个钟来测量走过每一段距离所需要的时间。那时认为长度是可以量的，因为尺子可以来回挪动，长度的测量是没有问题的。

然后你用你的钟来计量时间，用尺子来量长度，看一个作惯性运动的物体是不是在相同的时间段走过相同的距离。如果是的话，就说明你的钟是好的。这个钟是“好钟”。

后来又有人发挥了一下，不过基本仍是欧拉的那种思路。他们认为，一个好的钟应该使运动显得简单。什么意思呢？就是说“好钟”计量的时间应该使能量守恒定律成立，电磁学定律等也成立。但是什么叫运动简单，这个事情也很难说。欧拉他们对这个问题提供了一个思路，很具有启发性，但我觉得这个问题还没有很好解决。

相对论中如何确定“同时”

现在我们来看一看爱因斯坦怎样来校准不同地点的钟。在A、B两点，分别有一个钟，A钟与B钟。怎么校准这两个钟呢？图10-1左边是空间图，右边是时空图，在空间图中A钟发出光信号到B钟，B钟上有一个镜子把光信号反射回去。在时空图中，横坐标是A、B两个钟的空间距离，纵坐标是时间。钟A在 t_A 时刻发一个光信号到B钟，B钟在时刻 t_B 接收到，同时有一个镜子将它反射回去，信号在 t'_A 回到A钟。爱因斯坦规定光信号从A走到B的时间，与从B走到A的时间相等，即 $t_B - t_A = t'_A - t_B$ 。移项之后 t_B 就应该等于 $\frac{t_A + t'_A}{2}$ ，于是他就把 t_A 与 t'_A 的中点 \tilde{t}_A 定义

2

为跟B钟的 t_B 同时的那个时刻。就用这个办法来把A、B两个钟对好。其前提就是规定了光从A走到B的时间和从B走到A的时间相等。

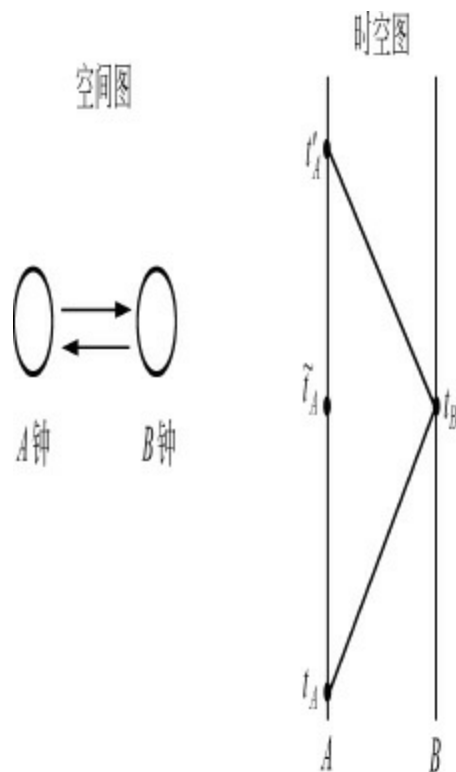


图10-1 惯性系中异地时钟的校准

爱因斯坦假定：“同时”可以传递

接着爱因斯坦又说：我假定同步性的这个定义是无矛盾的。

第一，我从A发射光信号到B再反射回来，和从B发射光信号到A再反射回去是一样的，用这两种方式定义的“同时”是一样的，不会有矛盾。

第二，假如有好几个相同的钟，放在不同地点，你用这种方法把A钟和B钟对好，再把B钟和C钟对好，那么C、A两个钟就自然对好了，不会有矛盾。

在爱因斯坦创立相对论的第一篇论文中，他就谈到了上述内容。这里特别值得注意的是第二条假定，我再重复一下。这条假定是说有三个钟，A钟、B钟、C钟，A钟发一个光信号到B钟反射回来，把这两个钟对好。B钟再发射一个光信号到C钟再反射回去，把B和C对好。那么C、A这两个钟就自然对好了，不会出现矛盾。

这是爱因斯坦在论文当中讲的，他的这段叙述在相对论发表之后很长一段时间没有引起大家的注意，大家觉得当然是这样，他讲的都对。大家都觉得没有问题。后来发现还真有问题。但爱因斯坦本人并没有说错，他说的上述内容都对，因为他当时是在惯性系中讨论问题。后来发现在弯曲的时空中，或者在平直时空的非惯性系中，“假设二”还真的会出问题，不过在惯性系中没有问题。

朗道：“同时”不一定能传递

苏联著名物理学家朗道提出怀疑，他说A钟的时刻和B钟的时刻对好，B钟的时刻和C钟的时刻对好，A、C就自然对好了。真是这样吗？在任何情况下都对吗？他在非惯性系或弯曲时空进行探讨，他用弯曲的时空坐标来研究，比如说一个转动的圆盘。在一个转动的圆盘上，你放三个钟，盘如果不转，那是惯性系，爱因斯坦说的内容都对。盘要一转，就不是惯性系了，你把A钟和B钟对好，把B钟和C钟对好，你就会发现A钟和C钟并没有对好。居然还有这种事，A钟和B钟对好了，B钟和C钟对好了，A、C两个钟居然就没对上。

怎样才能对上呢？朗道说你的时间轴，必须跟空间轴垂直。转盘上时间轴和空间轴没有完全垂直，有一个倾斜角度。我们通常看不见时间轴，你如果按相对论，把时间轴和空间轴都画出来，就会发现在转盘情况下，它们不垂直了。如果垂直的话就一定能够把钟对好。

你看A、B、C三个钟，A发射光信号到B，然后反射回来，对准A、B两个钟。然后B钟再与C钟对好，C呢再来对A。对完以后，朗道就发现A钟和B钟对好了，B钟和C钟对好了，C钟再跟A钟一对，它没有对到A钟原来那个时刻 t_A （图10-2），而对到了另一个时刻 t'_A 。什么情况下才能对回这个 t_A 呢？就是时间轴必须跟空间轴垂直才行。这是朗道和栗弗席兹著的《场论》上的内容。你看这个结论，真是非常奇怪。

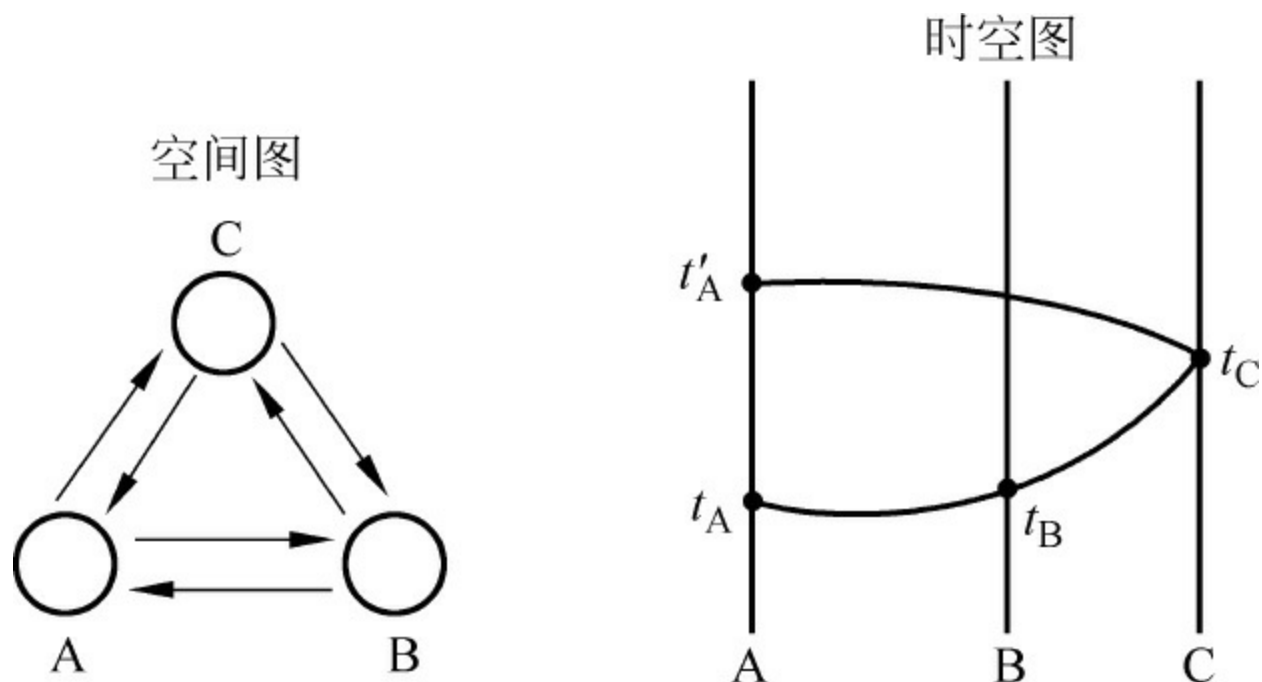


图10-2 同时的传递性

时间的定义与热力学定律有关吗？

我刚读研究生的时候，对这个问题特别感兴趣，因为刘辽先生在给我们讲广义相对论的时候讲到了这个问题。竟然还有这种事情。A和B对好，B和C对好，A和C两个却对不好。真令人匪夷所思。我当时想，物理学中还有没有别的东西是这样的？想来想去，我猜想热力学第零定律是不是和这个有点关联。

热力学第零定律说什么呢？学物理的人都知道，这条定理说：A、B两个系统达到热平衡，B、C两个系统也达到热平衡以后，A、C两个系统就自动地达到了热平衡。这条定律就是说“热平衡是可以传递的”。

朗道把他们的研究内容叫做“同时的传递性”，我就怀疑“同时的传递性”和热力学第零定律之间是不是有关系，就在研究黑洞的同时，不时地考虑这个事情，并真的动手去做了一些证明。

“钟速同步”的传递性

我在对钟这个问题上有一个小的创造。当时我的猜测是，如果在一

个参考系中热力学第零定律是正确的话，那么在该参考系中钟就能够对好，如果热力学第零定律不正确的话，钟就对不好，反过来也应该这样，如果在一个参考系中对钟对不好，那么热力学第零定律在这个参考系中就不正确。

但是我做这个证明之时，发现把时刻对好，这个要求太高了。第零定律只需要把三个钟的速率对好就行，不必要求把时刻也对好。我不要三个钟的时刻能对好，只要求这三个钟的快慢能对好：如果A钟走得跟B一样快，B钟跟C钟走得一样快，那么A、C两个钟就一样快，我称这个性质为“钟速同步的传递性”，并给出了“钟速同步具有传递性”的条件，这个条件是我首先给出来的。这个工作已经肯定是正确的了。感兴趣的读者可参看我与刘文彪教授合著的《广义相对论基础》，（清华大学出版社.2010）；或拙著《黑洞与弯曲的时空》（中国科技大学出版社.2013）。

后来，梁灿彬、高思杰等在他们的书和文章里做了推广和发展。表明这是一个新的对钟等级，跟朗道提出的不一样。我认为我给出的这个条件是和热力学第零定律等价的，我已在国内外的杂志上发表过几篇文章，当然这个讨论还没有做完。因为热力学在弯曲时空里怎么讨论，现在觉得很困难。

所谓钟的快慢，就像图10-3中线段的长短，如果这一段 $t_{A2}-t_{A1}=t_{B2}-t_{B1}$ ， $t_{B2}-t_{B1}=t_{C2}-t_{C1}$ ，是不是转回来的这一段 $t'_{A2}-t'_{A1}$ 和这一段 $t_{A2}-t_{A1}$ 是一样的。如果是一样的，就满足了“钟速同步具有传递性”条件，热力学第零定律就会成立。比如说匀速转动的圆盘，虽然“同时不具有传递性”，不能把A、B、C三个钟的时刻对好，但却能把这三个钟的钟速对好，也就是说“钟速同步具有传递性”。于是热力学第零定律就在匀速转动的圆盘上成立，如果圆盘不是匀速转动的，一般就不行了，这个条件就满足不了，钟速同步就不再具有传递性，因而热平衡也不会具有传递性。

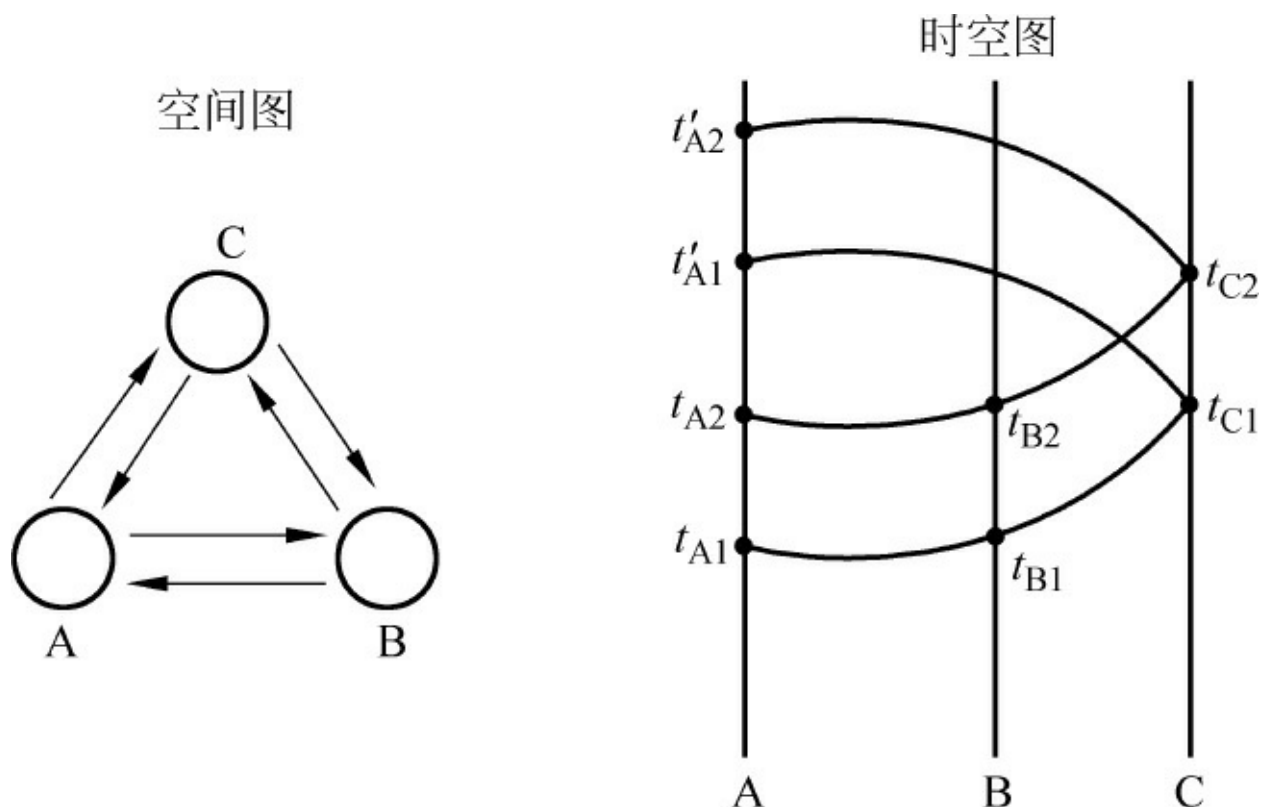


图10-3 钟速同步传递性的讨论

约定光速：寻找相等的时间段

我还要指出一件事：在“钟速同步具有传递性”的时空中，我们可以用这一对钟的方法，解决“相继时间段相等”的问题。大家再看一下图10-3，在此图中，把时间段 $t_{A2} - t_{A1}$ 与B、C两处的钟速（表现为时间段的长短）相继校准，即使得 $t_{B2} - t_{B1} = t_{A2} - t_{A1}$ ， $t_{C2} - t_{C1} = t_{B2} - t_{B1}$ 。再把C钟的钟速与A钟的钟速校准，即得时间段 $t'_{A2} - t'_{A1} = t_{C2} - t_{C1}$ 。这时，原来的 t_{A1} 没有对回到原处，而到达了 t'_{A1} ， t_{A2} 也未对回到原处，而到达了 t'_{A2} ，但满足了 $t'_{A2} - t'_{A1} = t_{A2} - t_{A1}$ ，这两个时间段相等了。如果我们调节 $t_{A2} - t_{A1}$ 的长短，使其经B、C两处再对回A处时， t'_{A1} 恰好与 t_{A2} 重合（图10-4），则线段 $t_{A2} - t_{A1}$ 与 $t'_{A2} - t'_{A1}$ 恰好连接，于是我们得到了相等的“相继时间段”。这正好解决了定义“相继时间段”相等的困难。

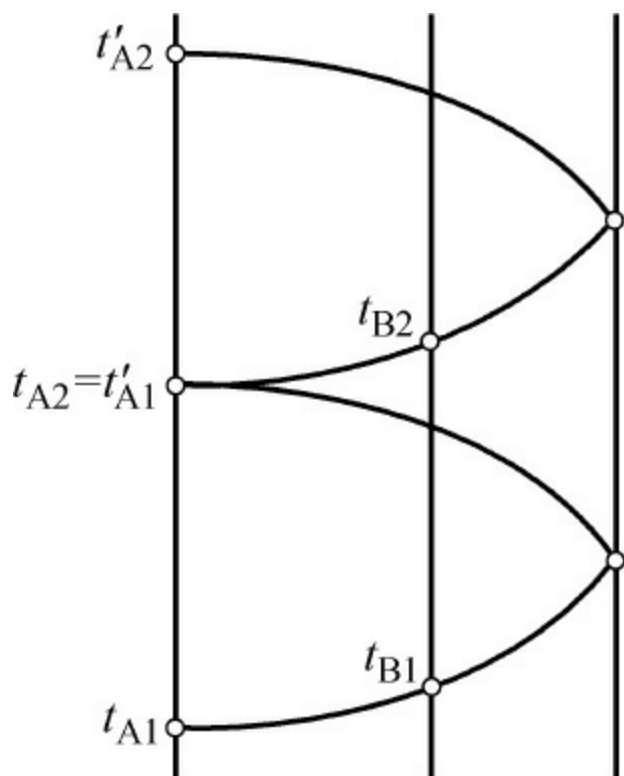


图10-4 相继时间段的相等（I）

事实上，操作可以更简化，只需要对一个时刻即可。如图10-5所示，让A钟的 t_{A1} 经过与B、C钟的时刻对好，回到A钟的时刻为 t_{A2} 。然后把 t_{A2} 再做一次与B、C钟校对，再次对回A钟的时刻为 t_{A3} ，显然 $t_{A3} - t_{A2}$ ，即图10-4中的 $t'_{A2} - t'_{A1}$ ，也就是说“钟速同步的传递性”可以保证 $t_{A3} - t_{A2} = t_{A2} - t_{A1}$ ，也就保证了“相继时间段”的相等。

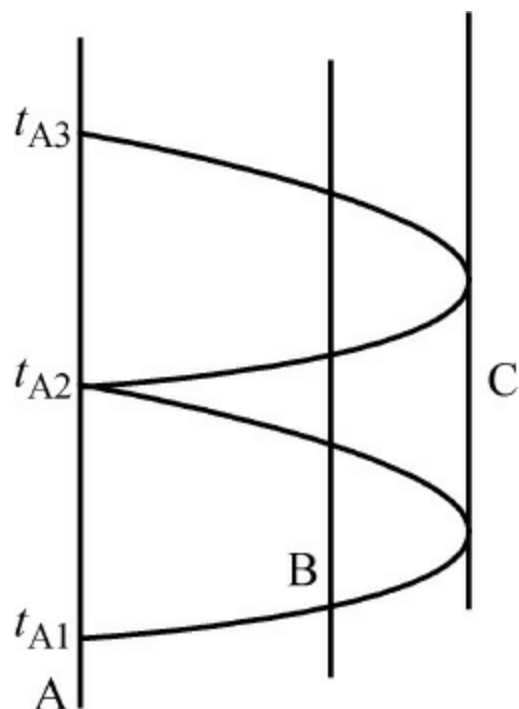


图10-5 相继时间段的相等（II）

时间的性质与热力学密切相关

注意，这一使“相继时间段”相等的操作，与异地时钟校准（包括钟速校准与时刻校准）的操作，用的是同一个“约定”（规定）：往返光速相等（即光速各向同性）。这样，我们就用同一个约定，解决了时间测量的两个问题。

对时间问题我们做了这些讨论，如果感兴趣，你们可以看我们写的书和论文。

一般人都知道热力学跟时间之间是有关系的。随便问一个搞物理的人，他都会告诉你热力学跟时间有关系，但大家主要指的是热力学第二定律。

热力学第二定律告诉我们时间是流逝的，有方向的。另外，搞物理的人还知道，热力学第一定律是能量守恒。能量守恒是和时间的均匀性有关系的。我们现在讨论了两个新问题，时间的无限性是不是和热力学第三定律有关；钟速同步的传递性是不是跟热力学第零定律有关。我认

为它们有关。

我曾经在普利高津研究所学习工作了两年多。我本来想把非平衡统计物理的问题搞成相对论性的，但这方面没有取得什么大的进展。但是后来，我对上述两个问题的研究倒是有所进展。虽然都是研究时间，但不是普利高津他们原来讨论、重视的，和热力学第二定律有关的东西，而是跟热力学第三定律和热力学第零定律有关的东西。

我今天的报告就到这儿，你们可能觉得很抽象。谁有什么问题，有没有？

问：您能不能讲一下，物理学家对“时间是什么”这个问题是怎样回答的？爱因斯坦是怎样认识这个问题的？谢谢。

答：其实，物理学家到现在为止，对时间是什么，没有给出很清楚的正面解答。但是你可以从一些物理学家的著作中看出他们的一些想法。

按照牛顿的观点，时间是一条均匀流动的河流，是可以脱离物质而存在的东西；爱因斯坦把时间和空间连到了一起，但是他的相对论也认为时空可以脱离物质而存在。在他的广义相对论中，有物质存在的时空是弯曲的，没有物质存在的时空是平直的。物质如果消失，时空就变平直，但依然存在。当然，在晚年，爱因斯坦的看法有转变，认为物质消失的话，时空也不应该存在。目前很多研究量子引力的人，采用了这一看法。

问：老师，假如时间是有开始和结束的，那么什么使时间开始呢？在时间开始之前是什么样的，因为在时间开始之前是没有时间的。

答：你说得对。在时间开始之前是没有时间的，所以就不存在“时间开始之前是什么样的问题”。至于“什么使时间开始”，有这方面的研究，但还没有得到清楚的答案。

时间是比空间更加值得让人深思的东西。然而，由于绝大多数物理分支都不考虑时间的流逝性，这样时间跟空间就相似了。所以很多搞哲

学的人就批评搞物理的人，说你们把“时间”空间化了。

后来我在自己的文章中答复他们，在物理学的大多数分支里，确实是这样，把时间与空间等同看待了，但是热力学第二定律是个例外。第二定律强调时间的流逝性和方向性。第二定律没有把时间空间化。而且，我们现在所知道的几种时间箭头，比如说宇宙学的时间箭头，热力学的时间箭头，心理学的时间箭头，还有什么其他的时间箭头，所有这些时间箭头根本上都起源于热力学第二定律，都可以归结为物理学的时间箭头。所以物理学的热力学分支是描述时间的流逝性的。而且所有其他的，对时间流逝性的论证，本质上都依赖于物理学的热力学分支，也就是说都是依赖热力学第二定律的。

我以前讲过，物理学中，除了热力学第二定律比较特殊以外，还有就是广义相对论比较特殊。因为其他理论都认为时空是平的，物质的存在对时空没有影响。只有广义相对论认为时空是弯曲的，物质的存在对时空的弯曲程度有影响，时空的弯曲也会反作用于运动的物质。

还有什么问题吗？我想今天这一讲，可能给大家提供了很多值得思考的东西。愿意思考的人，可能会觉得比较有趣。

插页诗句的注释与随想

1. 读赵嘏诗“闻笛”

唐代赵嘏（gǔ）的诗“闻笛”，优美高雅令人赞叹。而且该诗文字朴素，易于欣赏。诗曰：

谁家吹笛画楼中，
断续声随断续风。
响遏行云横碧落，
清和冷月到帘栊。
兴来三弄有桓子，
赋就一篇怀马融。
曲罢不知人在否，
余音嘹亮尚飘空。

由于“全唐诗”中无此诗，有人怀疑它不是赵嘏所作。然而，不管作者是谁，首先应该肯定，它是一首难得的好诗。

本书中那些对科学做过重大贡献的人，就像一个个吹笛的演奏家，时而吹出影响历史进程“响遏行云”的卓越成就，时而吹出增进人类幸福的美妙和声。他们自身的生活，有时曲折、悲壮，有时快乐、优雅，也像一曲曲跌宕起伏、婉转动人的笛音。因此我将赵嘏的这几句诗放在扉页，以飨读者。

诗中“碧落”二字是指道教主张的天的最高层，此句是说，美妙的音响把天上的行云都遏（è）止了，使它们横停在碧落天上，倾听这优美的笛声。

“兴来三弄有桓子”中的桓（huán）子是指东晋时的桓伊，他善于吹

笛，曾任淮南太守，在淝水之战中立过功。一次，在江边遇到书法家王羲之的儿子王徽之，徽之请他吹一曲笛子给自己听，桓伊越吹越起劲，一连吹了三遍。王羲之的几个儿子书法都很好，最好的是小儿子王献之。不过他们都不及其父。

“赋就一篇怀马融”。马融是东汉的大学问家，对《尚书》和《汉书》都颇有研究，他是伏波将军马援的姪孙，马续的弟弟。马续应班昭之邀，为《汉书》写了“天文志”。马融后来随班昭学《汉书》，学问大增。马融曾写过一篇“笛赋”，很有名。此句诗写作者听到如此优美的笛声，不由得想起马融的“笛赋”，自己也想写一篇类似的作品来表达感受。

2. 读王阳明诗“月夜二首”

这两首诗的标题下面有王阳明（即王守仁）自注：“与诸生歌于天泉桥”。据王的学生记载，那一年的中秋之夜，王阳明在天泉桥讲学，饮酒、奏乐、唱歌、吟诗、讲演、探讨。这样的讲学场面，真是别具一格，令人深思，令人向往。王即兴写了这两首诗。第一首是

万里中秋月正晴，
四山云霭忽然生。
须臾浊雾随风散，
依旧青天此月明。
肯信良知原不昧，
从他外物岂能撓。
老夫今夜狂歌发，
化作钧天满太清。

王阳明是唯心主义哲学家，认为“心”，“良知”是最根本的东

西。“霭（ǎi）”指云雾，“昧（mèi）”是灰暗的意思，“撖（yīng）”是触犯、扰乱的意思。“钧天”是天上的音乐，“太清”即太空、天空。前四句非常美好，末尾两句最有气魄，表现了王阳明自信、狂放的性格。

第二首是

处处中秋此月明，
不知何处亦群英。
须怜绝学经千载，
莫负男儿过一生。
影响尚疑朱仲晦，
支离羞作郑康成。
铿然舍瑟春风里，
点也虽狂得我情。

谈谈我对此诗的理解，不一定正确。

前两句的意思是，处处中秋都有此明月，但不知何处也像我们这里一样，聚集着才华横溢的青年。诗中充满了自信、自豪，感觉自己和自己的学子都是世间才俊。第三、四句是勉励年青人：应珍视这从远古流传、发展起来的学问，男子汉不要辜负了自己的一生。

第五、六句中的朱仲晦即南宋的理学家朱熹，郑康成即东汉的大学问家郑玄。朱熹主张“格物致知”，“格物穷理”。所谓格，就是去“想”，去“感通”。王阳明曾和朋友一起，按照朱熹的主张，坐在竹子面前“格”竹子，“格”了三天后那位朋友受不了了，退了下去。王阳明格了七天，什么也没有格出来，还差一点休克了。此后王阳明对朱的观点有了怀疑，所以“影响尚疑朱仲晦”。

郑玄是马融的学生，对“尚书”有很深的研究。“尚书”自古就有“今文《尚书》”和“古文《尚书》”两个版本，历史上一直存在两派学者的

争论。郑玄写了“尚书注”，大大推进了“古文《尚书》”的研究、教学与宣传，不过他的“尚书注”有许多观点不同于他的老师马融等老一辈学者，还有一些错误，后来遭到不少人批评。大概王阳明对他评价不太高，觉得他的理论有点破碎，所以“支离羞作郑康成”。

最后两句引用的是《论语·先进篇》中曾点言志的典故，曾点又名曾皙（xī），是曾参的父亲。一次，他和子路，冉有，公西华一起陪老师孔子坐。孔子让他们谈谈自己的志向，其他三人谈的志向都比较具体，谈话时曾点一直在弹瑟。最后孔子让他也谈一谈，他铿地一声放下瑟，说自己向往的就是在暮春三月，陪着亲友和孩子，在河边洗澡，在祭天求雨的舞雩（yú）台上吹风，然后唱着歌走回来。孔子十分赞扬这超脱、潇洒的精神境界，表示“吾与点也”，我与点的观点一样。王阳明认为这种自信，狂放，轻视物质欲望，追求精神高雅的志向与自己完全一样。

此诗的前四句最好，极有气魄，极有抱负。

王阳明是唯心主义的哲学家，提倡主观唯心主义。我不赞同他的哲学思想，但十分钦佩他的为人和治学精神。王阳明敢于批判学术权威，并独树一帜地构建新的学术思想，创建新的学派。

王阳明不是一个只会做学问的人，他关心世事，为官清廉，能文能武。不仅把地方治理得不错，打仗也很在行。无论是镇压人民起义还是平定藩王叛乱，他都干得很利索。

今天来看，王阳明做的事不一定都对；但从历史的角度看，他称得上是一位伟人。世界上没有完人，只有伟人，而伟人是可以通过奋斗来达到的。

王阳明的这两首优美、豪放的诗，充分阐释和表达了他的治学态度、奋斗精神，以及他对年轻人的勉励。本书中那些做出成就的科学伟人，都具有与王阳明类似的批判精神和奋斗精神。我希望借此机会，把王阳明的这种精神和他对后来学子的勉励介绍给读者。

主要参考书目

科普类

- [1] 爱因斯坦A. 狭义与广义相对论浅说 [M]. 杨润殷, 译. 上海: 上海科学技术出版社, 1964.
- [2] 霍金S W. 时间简史 [M]. 许明贤, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1994.
- [3] 彭罗斯R. 皇帝新脑 [M]. 许明贤, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1994.
- [4] 霍金S W. 霍金讲演录 [M]. 杜欣欣, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1994.
- [5] 霍金S W., 彭罗斯R. 时空本性 [M]. 杜欣欣, 吴忠超, 译.长沙: 湖南科学技术出版社, 1996.
- [6] 陶宏. 每月之星 [M]. 上海: 开明书店, 1949.
- [7] 伏龙卓夫·维略明诺夫. 宇宙 [M]. 郑文光, 译. 北京: 中国青年出版社, 1958.
- [8] 王允然, 褚耀泉. 从牛顿定律到爱因斯坦相对论 [M]. 北京: 科学出版社, 1982.
- [9] 卢米涅J-P. 黑洞 [M]. 卢炬甫译.长沙: 湖南科学技术出版社, 1997.
- [10] 索恩K S.黑洞与时间弯曲 [M]. 李泳, 译.长沙: 湖南科学技术出版社, 2000.
- [11] 吴国盛. 时间的观念 [M]. 北京: 中国社会科学出版社, 1996.
- [12] Novikov I. The River of time. 中译本. 时间之河 [M]. 诺维科夫. 吴王杰, 陆雪莹, 译. 上海: 上海科学技术出版社, 2001.

[13] Davies P C W. About time. 中译本. 关于时间 [M]. 保罗·戴维斯. 崔存明, 译. 长春: 吉林人民出版社, 2002.

[14] Poincare H. Science and Hypothesis. London: Walter Scott Publishing, 1905. 89-110.123-139. 中译本. 科学与假设 [M]. 彭加勒. 李醒民, 译. 北京: 商务印书馆, 2006.

[15] 彭加勒. 科学的价值 [M]. 李醒民, 译. 北京: 商务印书馆, 2007.

[16] 赵峥. 探求上帝的秘密 [M]. 北京: 北京师范大学出版社, 2009.

[17] 赵峥. 物理学与人类文明十六讲 [M]. 北京: 高等教育出版社, 2008.

[18] 郑庆璋, 崔世治. 相对论与时空 [M]. 太原: 山西科学技术出版社, 1998.

[19] 邓乃平. 懂一点相对论 [M]. 北京: 中国青年出版社, 1979.

[20] 陈应天. 相对论时空 [M]. 庆承瑞, 译. 上海: 上海科技教育出版社, 2008.

[21] 吴鑫基, 温学诗. 现代天文学十五讲 [M]. 北京: 北京大学出版社, 2005.

[22] 陆琰. 宇宙 [M]. 长沙: 湖南科学技术出版社, 1994.

[23] 刘学富, 李志安. 太阳系新探 [M]. 北京: 地震出版社, 1999.

[24] S. 温伯格. 最初三分钟 [M]. 冼鼎钧, 译. 北京: 科学出版社, 1981.

[25] 赵峥. 相对论百问 [M]. 北京: 北京师范大学出版社, 2010.

[26] 郭中一. 科学, 从好奇开始 [M]. 台北: 文经社, 2005.

[27] 倪光炯, 王炎森. 文科物理 [M]. 北京: 高等教育出版社, 2005.

[28] 倪光炯, 王炎森, 钱景华, 方小敏. 改变世界的物理学 [M]. 上海: 复旦大学出版社, 1998.

历史与科学史

[29] 郭奕玲, 沈慧君. 物理学史 [M]. 2版. 北京: 清华大学出版社, 2005.

[30] 吴国盛. 科学的历程 [M]. 长沙: 湖南科学技术出版社, 1995.

[31] Pais A. The science and the life of Albert Einstein. Oxford: Oxford Univ. Press 1982. 中译本. 爱因斯坦传 [M]. 方在庆, 李勇, 等, 译. 北京: 商务印书馆, 2006.

[32] 罗伯特·容克. 比一千个太阳还亮 [M]. 北京: 原子能出版社, 1966.

[33] 艾芙·居里. 居里夫人传 [M]. 左明彻, 译. 北京: 商务印书馆, 1981.

[34] L. Fermi. 原子在我家中 [M]. 何芬奇, 译. 北京: 科学出版社, 1979.

[35] 王自华, 桂起权. 海森伯传 [M]. 长春: 长春出版社, 1999.

[36] R. P.费曼. 爱开玩笑的科学家费曼 [M]. 吴丹迪, 吴慧芳, 黄涛, 译. 北京: 科学出版社, 1989.

[37] 乔治·伽莫夫.物理学发展史 [M].高士圻, 译.北京: 商务印书馆, 1981.

[38] 董光壁, 田昆玉. 世界物理学史 [M]. 长春: 吉林教育出版社, 1994.

[39] 范文澜. 中国通史简编 [M]. 北京: 人民出版社, 1965.

- [40] 周一良, 吴于廑. 世界通史 [M]. 北京: 人民出版社, 1973.
- [41] 刘家和. 古代中国与世界 [M]. 武汉: 武汉出版社, 1995.
- [42] 樊树志. 国史十六讲 [M]. 北京: 中华书局, 2006.
- [43] 李少林. 宋元文化大观 [M]. 呼和浩特: 内蒙古人民出版社, 2006.

科学著作与教材

- [44] 爱因斯坦A. 相对论的意义 [M]. 李灏, 译. 北京: 科学出版社, 1961.
- [45] Einstein A et al. The Principle of Relativity. Dover: Dover Publications. 1923. 中译本. 相对论原理 [M]. 爱因斯坦A. 赵志田, 刘一贯, 孟昭英, 译. 北京: 科学出版社, 1980.
- [46] Newton. I, Mathematical Principles of Natural Philosophy [M], Cambridge: Cambridge University Press, 1934.
- [47] 朗道, 栗弗席兹. 邹振隆, 校. 场论 [M]. 8版. 鲁欣, 任朗, 袁炳南, 译. 北京: 高等教育出版社, 2012.
- [48] S. 温伯格. 引力论和宇宙论 [M]. 邹振隆, 张历宁, 等, 译. 北京: 科学出版社, 1980.
- [49] 刘辽, 赵峥. 广义相对论 [M]. 2版. 北京: 高等教育出版社, 2004.
- [50] 梁灿彬, 周彬. 微分几何入门与广义相对论 [M]. 2版. 北京: 科学出版社, 2006.
- [51] 俞允强. 广义相对论引论 [M]. 北京: 北京大学出版社, 1987.
- [52] 须重明, 吴雪君. 广义相对论与现代宇宙学 [M]. 南京: 南京师范大学出版社, 1999.

- [53] 王允然, 鲁菲尼R. 相对论天体物理的基本概念 [M] . 上海: 上海科学技术出版社, 1981.
- [54] 张元仲. 狭义相对论实验基础 [M] . 北京: 科学出版社, 1994.
- [55] 赵峥. 黑洞的热性质与时空奇异性 [M] . 北京: 北京师范大学出版社, 1999.
- [56] 刘辽, 赵峥, 田贵花, 张靖仪. 黑洞与时间的性质 [M] . 北京: 北京大学出版社, 2008.
- [57] 赵峥. 黑洞与弯曲的时空 [M] . 太原: 山西科学技术出版社, 2000.
- [58] 赵峥, 刘文彪. 广义相对论基础 [M] . 北京: 清华大学出版社, 2010.
- [59] Wald R M. General Relativity [M] . Chicago and London: The University of Chicago Press, 1984.
- [60] Hawking S W, Ellis G F R. The large scale structure of space-time [M] . Cambridge: Cambridge University Press, 1973.
- [61] Birrell N D, Davies P C W. Quantum Fields in Curved Space [M] . Cambridge: Cambridge University Press, 1982.
- [62] Miller A I. Albert Einstein's Special Theory of Relativity [M] . London: Addison-Wesley Publishing Company Inc, 198. 185-200.
- [63] Rindler W. Essential Relativity [M] . New York: Springer-Verlag, 1977.
- [64] Misner C W, Thorne K S, Wheeler J A. Gravitation [M] . San Francisco: Freeman W H Company, 1973.
- [65] 王永久. 经典黑洞与量子黑洞 [M] . 北京: 科学出版社, 2008.

- [66] 王永久.经典宇宙与量子宇宙 [M] .北京：科学出版社，2010.
- [67] 李政道．场论与粒子物理学 [M] ．北京：科学出版社，1980.
- [68] Prigogine I. From being to becoming. San Francisco: Freeman W H and Company. 1980.中译本．从存在到演化 [M] ．普里戈金．曾庆宏，严士健，马本堃，等，译．上海：上海科学技术出版社，1986.
- [69] 喀兴林．量子力学与原子世界 [M] ．太原：山西科学技术出版社，2000.
- [70] 曾谨言．量子力学导论 [M] .2版．北京：北京大学出版社，1998.
- [71] 裴寿镛．量子力学 [M] ．北京：高等教育出版社，2004.
- [72] 赵凯华，罗蔚茵．新概念物理教程：光学 [M] ．北京：高等教育出版社，2004.
- [73] 赵凯华，罗蔚茵．新概念物理教程：量子物理 [M] ．北京：高等教育出版社，2001.
- [74] 王允然，李淑娴．力学概论 [M] ．合肥：安徽科学技术出版社，1986.
- [75] 郭硕鸿．电动力学 [M] ．北京：高等教育出版社，1997.
- [76] 梁绍荣，管靖．基础物理学 [M] ．北京：高等教育出版社，2002.
- [77] 李鉴增，狄增如，赵峥．近代物理教程 [M] ．北京：北京师范大学出版社，2006.
- [78] 何香涛．观测宇宙学 [M] ．北京：科学出版社，2002.
- [79] 李宗伟．天体物理学 [M] ．北京：高等教育出版社，2000.
- [80] 胡中为，萧耐园，朱慈盛．天文学教程 [M] ．北京：高

等教育出版社，2003.

后 记

本书引用了陆埏先生介绍的关于宇宙膨胀的资料，裴寿镛教授提供的关于量子力学的资料，裴申先生和杨静老师提供的天文学资料，杨再石先生提供的宇宙一词的出处。彭秋和教授与李庆康博士对书稿提出了修改建议。刘文彪、朱建阳教授也提供了许多帮助。笔者对他们表示深切感谢。

北京师范大学研究生刘艳芳、本科生游莉莉、张宁、韩晶、陈陟陶花费大量时间，帮助笔者整理文稿，研究生鹿鹏举、谢丽璇也对笔者提供了帮助，笔者在此表示深切谢意。

清华大学出版社朱红莲、石磊、邹开颜等老师大力支持并协助这本演讲集的出版，作者在此深表感谢。



本书来源于作者长期以来在北京师范大学开设的一个科普讲座“从爱因斯坦到霍金的宇宙”。该讲座历时20余载，并在一些院校和单位举办过不同形式的讲座和公开课。重点介绍物理学和天文学领域的科普知识、科研前沿，以及科学发现的曲折历程，内容包括爱因斯坦与相对论、弯曲的时空、黑洞、宇宙的演化、量子论的创建与争论、原子弹与核能的和平利用、天文学的若干知识、对时间本质的探索等。

理解科学丛书 卢昌海科普著作

LITTLE HALLS
People and Stories
in the Temple of Science AND GRAND MASTERS

小楼与大师

科学殿堂的人和事

卢昌海◎著

理论的沉浮、智者的瑕瑜……

小楼里的大师、乱世中的学人……

美丽的科学史话、动人的探索故事……

清华大学出版社

作者简介

卢昌海，出生于杭州，本科就读于复旦大学物理系，毕业后赴美留学，于2000年获美国哥伦比亚大学物理学博士学位，目前旅居纽约。著有《那颗星星不咋星图上：寻找太阳系的疆界》、《太阳的故事》、《黎曼猜想漫谈》和《从奇点到虫洞：广义相对论专题选讲》，并曾在《中国青年报》、《数学文化》、《科幻世界》、《现代物理知识》、《中学生天地》、《科学画报》等报纸、杂志上发表几十篇科普及专业科普作品。





理解科学丛书

LITTLE HALLS
AND GRAND MASTERS
People and Stories
In the Temple of Science

小楼与大师

科学殿堂的人和事

卢昌海◎著

清华大学出版社
北京

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

小楼与大师：科学殿堂的人和事 / 卢昌海著. --北京：清华大学出版社，2014

（理解科学丛书）

ISBN 978-7-302-35965-4

I. ①小... II. ①卢... III. ①科学史-青少年读物 IV. ①G3-49

中国版本图书馆CIP数据核字（2014）第066035号

责任编辑：邹开颜

封面设计：蔡小波

插图：张京

责任校对：赵丽敏

责任印制：

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印 刷 者：

装 订 者：

经 销：全国新华书店

开 本：165mm×240mm

印 张：12.75

字 数：179千字

版 次：2014年6月第1版

印 次：2014年6月第1次印刷

印 数：1~0000

产品编号: 054823-01

谨以本书献给我的家人

自序

作为一位出版了四本书的作者，如果要用一句话来概括写书的感觉的话，那就是：写书比写文章累。这貌似是一句显而易见的大白话，对我这种在写作上有一定兴趣，甚至以写作为乐的人来说，却是一种只有经历过了才意识到的新感觉。这新感觉的起因也是一句显而易见的大白话，那就是：书比文章长。不过，这个“长”对我来说与其说是篇幅之长，不如说是指所费时间之长。因为在一本书的写作过程中，我得不断约束自己的阅读兴趣，把主要精力投注于单一主题。另一方面，我的写作速度又比较慢（或美其名曰“认真”），从而使得写作过程往往长到了对题材的兴趣将尽而书稿远未完成的程度。这时候，写书就变成了对恒心和毅力的考验，而我——很遗憾地——曾两度在这种考验面前失败过，致使《黎曼猜想漫谈》和《从奇点到虫洞》“烂尾”多年（对这一“丢人”事迹感兴趣的读者可参阅那两本书的后记），其“累”亦由此可见。

在这种感觉下，若有谁愿把我的文章汇集成书出版，让我既免除写书之累，又可得出书之乐，那对我来说简直就是“天上掉馅儿饼”的美事，几乎要生出一种“偷懒”的愧疚了。最近，这样的美事居然落在了我的头上——清华大学出版社愿意出版我的两篇文章合集，一本收录科学史方面的文章，一本收录科普方面的文章。

兴奋之下，我很快选好了篇目，但问题来了：一堆文章汇集在一起，以什么作为书名呢？当然，假如我是著名作者，这根本就不是问题，大可取名为《卢昌海科学史作品集》和《卢昌海科普作品集》。但对于明显不著名的我来说，就算不怕僭越地将自己的名字厚颜纳入书名，也只会成为“票房毒药”，因此必须另谋思路。读者可能会笑话我这

么小的事情都不能轻松搞定，其实非独我如此，像阿西莫夫（Isaac Asimov）那样的大牌作家也常常为书名发愁呢，以至于在文章合集*The Sun Shines Bright*的简介中感慨说，他几乎想用数字编号来作书名了——当然，他发愁的原因跟我是不同的，他那是因为作品实在太多，显而易见的书名几乎用遍了。

经过思考，为了让两本书略显对仗，我提议将科学史合集取名为《科学殿堂的人和事》，将科普合集取名为《科学殿堂的砖与瓦》。但编辑看了之后觉得这两个标题太平淡。于是我又绞尽脑汁想了半天，却没再想出什么点子来。无奈之下，我决定效仿阿西莫夫，他虽然也为书名发愁，点子可比我多多了，在*The Sun Shines Bright*的简介中做完了用数字编号作为书名的“白日梦”后，随即采用了一个颇有些取巧的办法，那就是从所汇集的文章中选取一篇的标题作为书名。现在您所看到的这两篇文章合集的书名——《小楼与大师：科学殿堂的人和事》和《因为星星在那里：科学殿堂的砖与瓦》——便也是如此而来。

关注我文章的读者或许注意到了，收录在这两本书中的某些文章是曾经在杂志或报纸上发表过的。不过，杂志和报纸大都有自己固定的风格，有时不免需要作者“削足适履”来契合之。因此，发表在杂志和报纸上的版本与我自己的版本相比大都存在一定的缺陷，比如经过编辑的改动，以及因字数所限作过删节等。此外，发表在杂志上的版本大都略去了注释及对人名和术语的英文标注等，这其中后者——即英文标注——或许并不重要，但前者——即注释——其实是颇为重要的，往往起着补充正文、澄清歧义等诸多作用。所有这些缺陷在此次汇集成书时都尽可能予以消除了。

与以前的四本书一样，这两本书也是非常接近原稿风格的，在个别细节上甚至有可能略胜于原稿，因为编辑订正的个别错别字由于未曾标

注，我未必能在阅读校样时一一察觉并在自己的版本上做出相应的订正。在尊重原稿这个最至关重要的特点上，我要再次对清华大学出版社表示感谢，感谢其对我作品及写作风格的长期——从出版第一本书至今已五年了，够得上用这个词了吧——信任和支持。

最后，希望读者们喜欢这两本新书。

目 录

[自序](#)

[第一部分 科学史](#)

[书林散笔：小楼与大师](#)

[书林散笔：陨落的前辈](#)

[泡利效应趣谈](#)

[让泡利敬重的三个半物理学家](#)

[纪念戈革——兼论对应原理、互补原理及EPR等](#)

[一、著作等身的翻译家](#)

[二、经历坎坷的独行侠](#)

[三、关于玻尔的历史地位](#)

[四、关于对应原理](#)

[五、关于互补原理](#)

[六、关于EPR争论](#)

[七、结语](#)

[玻尔的错误](#)

[一、引言](#)

[二、玻尔的第一次错误：BKS理论](#)

[三、BKS理论的放弃](#)

[四、玻尔的第二次错误：科学革命综合症](#)

[五、结语](#)

[希尔伯特与广义相对论场方程](#)

[一、引言](#)

[二、希尔伯特对物理学的兴趣](#)

[三、希尔伯特的《物理学基础》](#)

四、早期研究简述

五、校样风波

六、信件辨析

七、“借鉴”之争

八、尾声

他们为什么反相对论？

一、先声

二、德国的反相对论运动

三、苏联的反相对论运动

四、中国的反相对论运动

五、结语

从爱因斯坦妻子的“秘密贡献”谈起

希尔伯特第十问题漫谈

一、问题

二、算法

三、丢番图集

四、罗宾逊猜想

五、解决

附录：丢番图方程

黎曼猜想浅说

一

二

三

四

五

六

第二部分 其他

[科学的目](#)

[科学的方法](#)

[科学哲学讨论中的“大规模杀伤武器”](#)

[关于批评的资格](#)

[人名索引](#)

[术语索引](#)

[返回总目录](#)

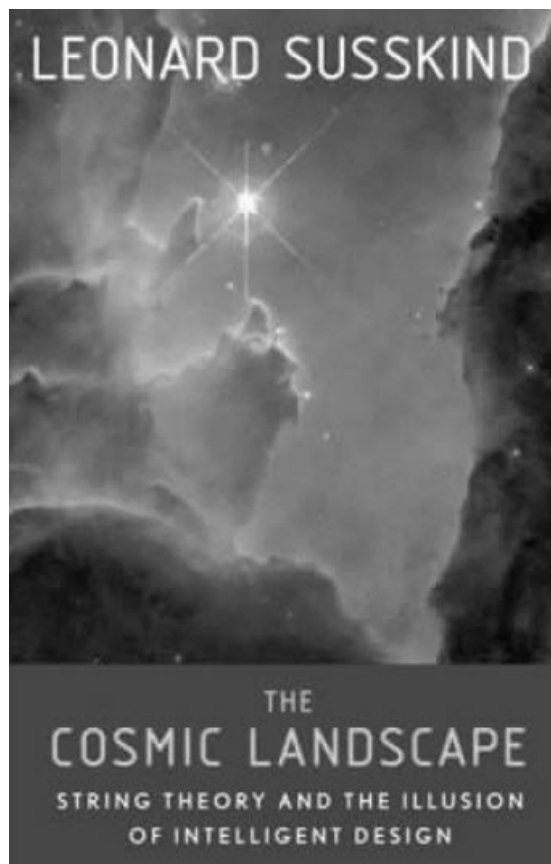
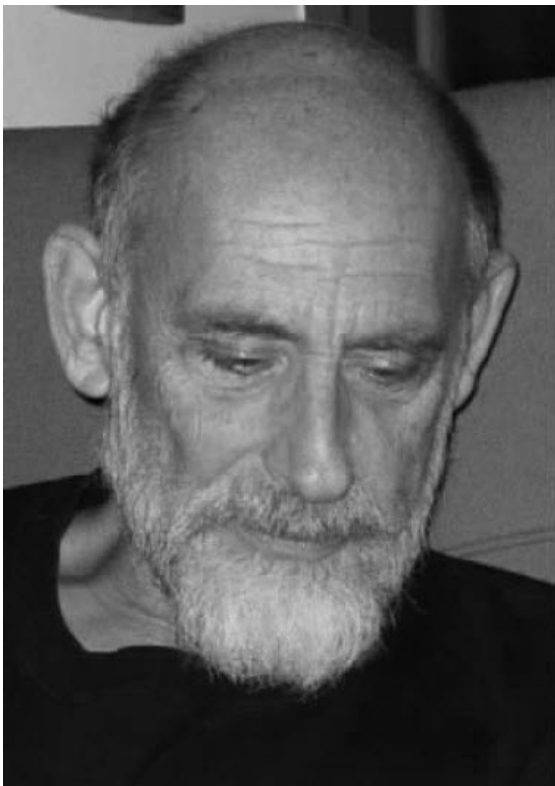
第一部分 科学史

书林散笔：小楼与大师

几星期前在书店闲逛时，发现美国物理学家苏士侃（Leonard Susskind）的*The Cosmic Landscape*一书被放在了特价区，惨遭贱卖（倒不是因为滞销，而是由于书的侧面不知怎的染上了一些黑色墨迹），于是毅然解囊买了一本。

苏士侃这本书是2005年底出版的，讲述的是他不久前做出的一项比较得意、且引起广泛兴趣及争议的工作。他这本书我以前曾买过一本送朋友，自己却未曾阅读，只记得书中所附的作者近照有点像《星球大战》（Star Wars）中的杜库（Count Dooku）——其实只是脸型和胡子有几分像，那是黑暗尊主（Dark Lord）的徒弟，是一个很厉害的坏蛋，曾砍去阿纳金·天行者（Anakin Skywalker）——也就是后来的黑武士达斯·维达（Darth Vader）——的右手（但最终还是被阿纳金·天行者所杀）。

苏士侃今年已经68岁，是斯坦福大学（Stanford University）的理论物理学教授。他和获得2008年诺贝尔物理学奖的南部阳一郎一样，都是弦理论的创始人之一。不过与南部不同，苏士侃直到今天仍在与弦理论有关的领域中工作，并做出了不少有创意、且很能吸引公众注意力的工作——比如在黑洞与信息、弦景观（string landscape）与人择原理等方面的工作。不过，在这个以闲谈为主的“书林散笔”系列中，我们不谈太学术的话题，因此苏士侃的那些学术工作将留待今后再单独介绍。在本文中，我想讲述一段苏士侃在*The Cosmic Landscape*一书中提到的他本人亲身经历的有趣见闻。



苏士侃和他的*The Cosmic Landscape*

这段见闻发生在苏士侃的年轻时代——确切地说是发生在他前往纽约曼哈顿北部的叶史瓦大学（Yeshiva University）就任助理教授的那一天。那时苏士侃已从康奈尔大学（Cornell University）获得博士学位，并在加州大学伯克利分校（University of California, Berkeley）做完了一年的博士后工作。叶史瓦大学的那个职位是他当时手头的唯一职位。据他在*The Cosmic Landscape*中所述，这段经历发生在1967年，不过苏士侃的简历所列的叶史瓦大学任职时间却为1966—1970年，因此在时间上不排除有一年的误差。叶史瓦大学是一所私立的犹太人大学，创建于1886年，据说是美国最早的犹太人高等学府，2008年的全美排名是50。叶史瓦大学离我就读过的哥伦比亚大学（Columbia University）不远（在哥伦比亚大学以北5~6公里处），那里早年有很多犹太中产家庭居

住，但后来逐渐变成了拉丁裔移民的聚居区。

曼哈顿是一个寸土寸金的地方，在曼哈顿的大学中，除哥伦比亚大学有一个很小的校园外，其余大都只有一些被街道分隔开的建筑。苏士侃刚从康奈尔及伯克利那样环境优雅的校园来到这里，视觉上的反差是可想而知的。我以前住在曼哈顿的时候，曾多次路过曼哈顿南部的纽约大学（New York University），该校的建筑就分散在若干个街区。为了便于识别，那些建筑统一悬挂了青紫色的校旗，每次看到那些校旗，让我想起的不是大学，而是《水浒传》里那些小店门口迎风招展的酒旗。

不过对苏士侃来说，校园环境的好坏还在其次，因为更坏的事情很快就出现在了的面前。那天他来到叶史瓦大学的校区，在路人的指引下，找到了物理系。那是一间很小很昏暗的屋子，屋子里有一个大书架，书架上放满了大部头的书本，但不是物理书，而是希伯来文的古书。屋子里还有一把椅子，椅子上坐着一个胡子灰白的家伙，正翻看着一本古书。

苏士侃的到来使宾主双方展开了历时一分钟的坦诚而富有建设性的会谈，那人向客人详细介绍了物理系的状况：

- （1）他这间小屋就是物理系。
- （2）他就是物理系的系主任。
- （3）物理系有且仅有一位教授。
- （4）那位教授就是他自己。
- （5）他从未听说物理系招了助理教授。

在美国像这种“一人吃饱，全系不饿”的物理系其实并不罕见，我哥伦比亚大学的一位师兄就一度在那样一个物理系里工作过——当然，那意味着当过系主任。听完系主任的热情介绍，苏士侃的心沉到了海底——他还没上岗，看来就先要失业了。苏士侃当时虽然才26岁，却不仅已早早结了婚，而且还有两个小孩，这拖妻带幼的谋生，那是相当的不容易。但情势如此，夫复何言？他只得离开那皮包公司般的物理系，退回了街上。

幸运的是，在街上他碰到了一个人，于是他向那人叙述了自己的遭遇，那人听了哈哈大笑，说：“你想去的大概是研究生院的物理系，而不是本科的物理系吧。”笑罢，他告诉了苏士侃研究生院的地点。

苏士侃大喜过望，赶紧找到新地址。可到了那里——几个街区外的一个街口——一看，满眼皆是破烂店铺，有的甚至已经废弃，苏士侃再度失望。难道连朋友也会忽悠自己？他不死心地又在那街口转了一圈，结果突然在一家早已废弃的犹太婚宴店旁发现了一块小牌子，上面写着：**Belfer Graduate School**（贝尔夫研究生院）。那牌子指向一串楼梯，那楼梯污垢不堪，地毯也早已破旧，楼道里没有书香，却弥漫着食物的气味。看来这地方并不比那本科生的物理系强。

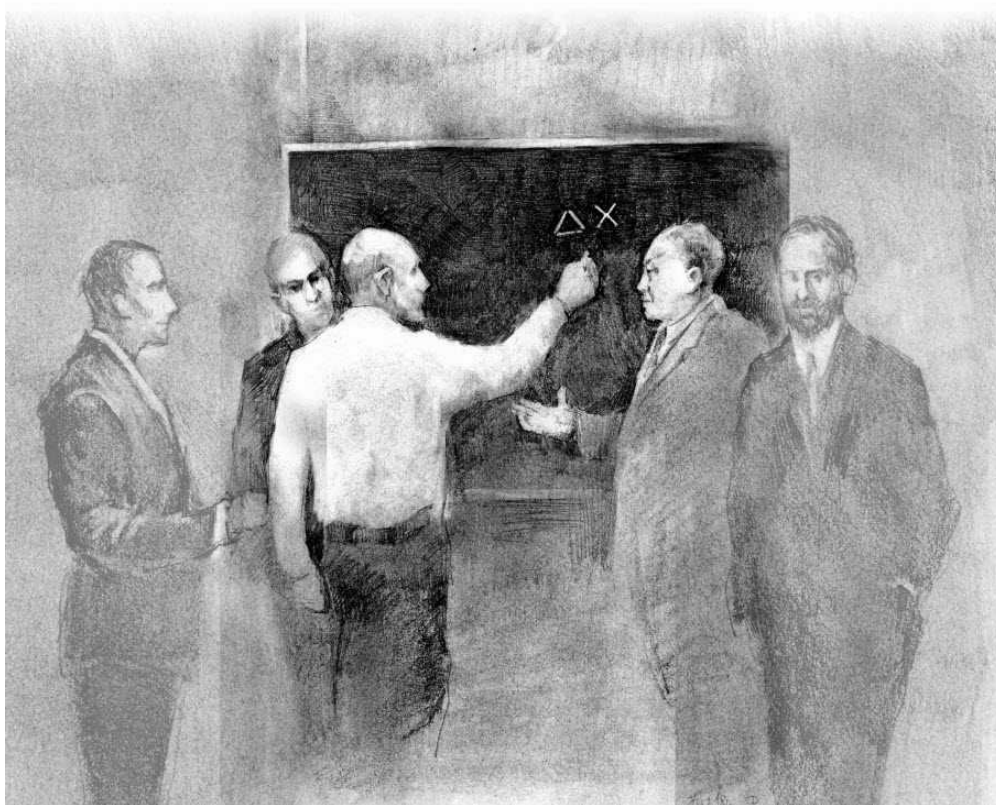
苏士侃硬着头皮登上楼梯，上面是一个很大的厅，估计是那家犹太婚宴店昔日所用的舞池。大厅的四周约有二十来间办公室。看来这破烂小楼上的废弃舞池就是整个的研究生院了。不过，这时苏士侃看到了一件让他松了一口气的东西：黑板。在经历了方才的连番遭遇后，这小小的黑板在苏士侃的眼里变得亲切无比，简直就像救命稻草——用他自己的话说：“黑板意味着物理学家”。

但是，最让苏士侃振奋，并且让所有阴暗的氛围一扫而空的，则是

人——几个正在黑板旁讨论问题的人。这是苏士侃这段见闻中最富戏剧性的地方，它大大出乎苏士侃的意料——当然，也完全出乎我的意料。那几个参加讨论的人苏士侃全都认识，他们之中包括这样几位：

- 芬克尔斯坦（David Finkelstein）——“爱丁顿-芬克尔斯坦坐标”（Eddington-Finkelstein coordinates）中的那个芬克尔斯坦。
- 阿哈罗诺夫（Yakir Aharonov）——“阿哈罗诺夫-玻姆效应”（Aharonov-Bohm effects）中的那个阿哈罗诺夫。
- 勒波维茨（Joel Lebowitz）——“勒波维茨不等式”（Lebowitz inequalities）中的那个勒波维茨。
- 彭罗斯（Roger Penrose）——“霍金-彭罗斯奇点定理”（Hawking-Penrose singularity theorems）中的那个彭罗斯。
- 狄拉克（Paul Dirac）——这位老同志还需要介绍吗？

在这几位当中，芬克尔斯坦、阿哈罗诺夫、勒波维茨当时都在叶史瓦大学任教，彭罗斯和狄拉克则是访问学者。别看这地方寒碜，狄拉克早在1964年就莅临视察过，并且还发表过重要讲话，他的讲话内容后来被整理成了*Lectures on Quantum Mechanics*，是有约束量子理论的经典著作（该书有中译本：《狄拉克量子力学演讲集》，由科学出版社出版）。



这破旧小楼上的那几位物理学家是苏士侃在单一地点所见过的最杰出的一群物理学家，他们的出现彻底扭转了他那一天的心情。只要有这些人在，哪里不能是物理系呢？那一天，那几位物理学家讨论的话题是真空能（vacuum energy），这个话题从此成为了苏士侃毕生探究的课题之一，这种探究最终导致了他这本*The Cosmic Landscape*的问世。

我记得几年前有很多人在讨论“大楼”与“大师”的问题。我想，苏士侃这段小经历对于那个话题是一个很好的参考。其实，在中国自己的教育史上，就有过一所没有半栋大楼，却有很多大师的大学：国立西南联合大学。那所大学只存在了短短八年，却在中国教育史上留下了永不磨灭的一页。正所谓“山不在高，有仙则名，水不在深，有龙则灵”。如果让中国的学人评选一所中国最杰出的大学，我想只要国立西南联合大学是其中一个选项，它就一定会夺冠。

二零零八年十一月六日写于纽约

书林散笔：陨落的前辈

要是有这么一个地方，最好是靠海的地方，没有会议，没有斗争，也没有这么多的莫名其妙的麻烦事。带上几个学生，安安静静地搞项目搞研究，该多好。

束星北（1962年——时为极右派兼历史反革命分子）



束星北



《束星北档案》

几年前，我建过一个论坛，叫做繁星客栈，那里聚集了一些很不错的网友。有一天，一位网友转了一篇题为《一部浮夸的科学家传记——

评刘海军〈束星北档案〉》的文章，那是我第一次听说《束星北档案》这本书，而且有可能也是我第一次听说束星北这个人。

那篇文章我虽只是粗略看了看，却留了一个印象。我曾看过不少国内作者撰写的华人科学家传记，比如杨振宁、李政道、丁肇中、吴健雄、苏步青、谢希德、“三钱”等的传记，可以说无一例外是高、大、全式的歌功颂德。有这样的体验做后盾，虽未看过《束星北档案》，见有人从浮夸的角度进行批评，倒也不觉意外。后来有一次回国，在书店里看到了《束星北档案》，想起那篇评论，便没有购买。

此后又过了很长时间，我几乎已将那本书忘了，不想却在纽约的一家图书馆里看到了它，于是借了回家。看完之后，我不无诧异地发现这本书在我读过的华人科学家传记中几乎可算是最好的（当然，这要部分地归因于其他传记的过于乏善可陈）。这本书虽的确对束星北的学术水平作了少许外行及浮夸的评价^①，但重点并不在他的学术，而是在叙述他坎坷的人生经历。在全书30个章节（包括引言和尾声）中，与学术及教学有关的内容大都集中在前三章，比例极小。因此，尽管作者在讲述束星北的学术经历时，确实作出或引述过一些夸大其词的评价，但从内容比例上讲，这本书给人的真正印象并非是束星北是一个如何了不起的物理学家，而是他在政治漩涡中苦苦挣扎的佝偻背影，以及他可怜可叹的人生悲歌，这与其他那些传记是截然不同的。

从内容上讲，这本书最独特的地方，是它采用了大量的采访记录及档案资料，其中包括束星北本人所写的许多“思想汇报”。我之所以欣赏这本书，最主要的原因也在于此。事实上，在作者仰视传主及作者的学术外行性这些传统缺陷上，《束星北档案》与其他那些歌功颂德式的传记未必有很大差别，但该书采用的引述采访记录及档案资料的做法在很大程度上弥补和淡化了那种缺陷。因为那些来自不同人、不同时期、不

同视角的回忆资料大都很平实，其中有很多描述的是束星北的窘态（当然，外行人道听途说的过誉之词也是有的，但即便那样的回忆也并无刻意歌颂的意味，并且大都言明了是道听途说）。而束星北本人那“一把辛酸泪，满纸荒唐言”的“思想汇报”，不仅没半点高、大、全的模样，反而充满了自我羞辱。有这些扎实的史料作基础，这本传记无论对于了解束星北这个人，还是了解当年那个吃了人还恬不知耻地让被吃之人歌颂自己的新社会，都是难得的第一手资料⁽²⁾。

好了，现在回过头来说说束星北这个人。束星北出生于1907年，与那个让他受尽折磨的新社会有着共同的生日：10月1日，是中国物理界的一位前辈。束星北的求学经历相当奇特，美国学者胡大年在《爱因斯坦在中国》一书中曾有一节用了“游学四方的束星北”作为标题。束星北在1924—1931这八年的求学期间，曾辗转于杭州之江大学（其校址在现浙江大学之江校区）、山东齐鲁大学（其校址在现山东大学医学院）、美国贝克大学（Baker University）、美国加州大学旧金山分校（University of California, San Francisco）、德国汉诺威工业大学（University of Hanover）、英国爱丁堡大学（University of Edinburgh）、英国剑桥大学（University of Cambridge）及美国麻省理工学院（MIT），并先后在爱丁堡大学及麻省理工学院获得过硕士学位⁽³⁾。束星北的这种奇特的求学史，看来体现了一种躁动的性格，他的同时代人吴大猷先生曾在《早期中国物理发展之回忆》一书中评论说束星北在欧洲和美国跑来跑去，没有认真地待在哪个地方做研究，他的学生许良英也曾在一篇纪念文章中提到束星北缺乏专心致志的精神。束星北是一个有才之人，除物理外，在后来因生活所迫而改做的气象、化工乃至电器修理等工作上也都有不俗的表现，可惜却一生并无建树，究其原因，除不幸生活在一个特殊时代外，与他躁动的性格恐也不无关系。

束星北1931年回国后，曾在南京中央军官学校、暨南大学、浙江大学等地任教。在1952年的院系调整中，浙大遭到肢解，束星北选择了山东大学继续自己的教学生涯。1954年，束星北因公开反对马列主义统领一切的观点而遭批判，并被逐出物理系，开始了他后半生的漫长噩运。离开物理系后，在竺可桢与王淦昌的关照下，束星北在山东大学气象研究室转行研究气象动力学，并在一年多的时间里发表了十几篇论文。

但他这一短暂的庇护所也很快失守。1955年，气象研究室在“肃反”运动中被关闭，束星北夫妇遭到批斗和体罚，束星北一度萌生了自杀的念头。1956年，“鸣放”运动开始了，被这些风向迥异的政治运动搞得稀里糊涂的很多知识分子以为春天终于来临了，纷纷将前一阶段的苦水倒了出来，结果中了“引蛇出洞”的阳谋，在接踵而来的“反右”运动中被一网打尽，这其中也包括束星北。

在“鸣放”期间，束星北发表了一些我个人非常欣赏的观点。在旁人——包括很多高级知识分子——都只见树木不见森林的时候，束星北敏锐地看出了无论“肃反”还是“鸣放”，它们的主题虽截然不同，形式却如出一辙，全都充满了非理性和不守法的“人治”及“发泄”特点。1957年5月，束星北作了一次题为《用生命维护宪法的尊严》的报告，主张推进法制。束星北的这种冷静和理性的思维，直到今天仍是中国社会相当欠缺的，因为直到今天，迎合大众胃口的谎言依然能轻易而迅速地调动巨大的非理性力量。

可惜，越是冷静和理性的思维对专制的威胁就越大。因此尽管束星北在“鸣放”期间的行为相较于激进人士来说显得很克制，但“反右”运动一降临，他很快就被戴上了“极右派”和“历史反革命分子”等帽子，于1958年10月与1800名其他右派一起被押往山东夏庄强制劳动。他们的劳动结果便是今天的崂山水库（当时叫做月子口水库）。在那里，劳动条

件极其艰苦，工伤、自残和自杀时有发生；在那里，束星北吃尽了苦头，失尽了尊严。在1959年开始的那名为“自然灾害”，实为“大跃进”导致的三年人祸期间，他四处借贷，甚至不得已到田里去偷地瓜（结果被当场抓住）。



与此同时，他的所有子女都受株连，工作丢了，对象吹了，有的甚

至被劳教。在这种无可抗拒的力量面前，束星北终于“顿悟”到螳臂挡车是没有出路的，并开始递交一些歌颂时局、自我改造的思想小结。随着水库工程的完工，束星北被遣到青岛医学院继续改造，当年的同事曾这样回忆这位经过改造的社会主义“新人”：

束先生是拄着拐杖来校报到的，他和几年前我见的那个神采奕奕、侃侃而谈、高声大嗓的先生已判若两人。要不是保卫科李科长在前面引着，我无法相信站在面前的这个憔悴、浮肿、目光散淡的老人就是束星北。他头顶着蓝色带护耳的棉帽，双手支在拐上，背抵在墙上，好像不这样“夹”着，人随时就会倒下去。李科长当着物理教研组全体成员给他训话时，他就像个泥塑木雕，身子和眼睛好长时间也不动一下。

但即便在这样的境况下，束星北仍然做出了一件一鸣惊人之事。1961年，他在干杂活之余，帮青岛医学院修好了一台损坏已久的进口脑电图机，引起了轰动。在那之后，青岛乃至外地的很多医院都慕名请他修理仪器。虽然修理仪器在束星北眼里只是雕虫小技，但自己重新变得有用还是鼓起了他的勇气，他开始希望凭借自己的才华来向党和人民“赎罪”，以便“摘帽”。为此他废寝忘食地努力着，并递交了“摘帽”申请。他的努力得到了一定层级的表扬，他心中的期盼也因此而变得更为炽热。青岛医学院公布第一批“摘帽”名单的那天，他拄着拐杖满怀希望地前往会场，可惜“摘帽”名单中并没有他。宣布完名单后，可怜的老先生茫然失措，痴痴地站在会场里直至人群散尽。



苦苦等待“摘帽”的束星北夫妇

受到沉重打击的束星北并不清楚自己的问题究竟出在哪里？自己究竟怎样才能被“改造”好？他一方面更加努力，另一方面则以最诚恳的态度请大家给他提意见，甚至主动请求开一个针对他的评审会。有道是“精诚所至，金石为开”，在他的一再恳请下，“劳动人民”终于向他吐露了心里话。在评审会上，他听到了这样的批评：“你的书本知识有一些，志大才疏，但还高傲，你实际上还是个矮子”，“你老想从科研上找出路”，“你想单干，修仪器赚钱”。束星北的觉悟再低，到这时也基本体会到了“劳动人民”的立场。“劳动人民”虽然缺乏知识，但并不缺乏忌妒心，想要凭借“劳动人民”所不会的技能来“摘帽”，那只会更让人家觉得他高人一等。束星北再次“顿悟”，要想让“劳动人民”满意，不能当专家，而必须当孙子。于是他自请长期打扫医学院所有的茅房，以彻底改造自己。

束星北打扫的不是五星级公厕，而是混杂了痰迹、粪便乃至人体器官碎片的医学院茅房，并且在打扫的过程中有时得用手去抠被这些东西堵塞的大便池，其情形是如今的我们不易想象的。他一边打扫，一边不断地提交思想汇报，在汇报中他写道：“刷茅房就是具体地听党的话，具体地为人民服务，因此刷茅房也就意味着刷掉资产阶级的臭思想”，

他并且表示现在刷茅房还感觉到脏臭，今后的目标是要做到“完全自然，不感觉脏臭”。三个月后，他在思想小结中写道：“刷了三个月的茅房之后，越刷越起劲，越刷越愉快”。束星北所说的“越刷越愉快”倒不一定是谎话，因为干这样的活，“劳动人民”是不会找他麻烦的，这样干下去，说不定“摘帽”也会重新有望。人总是在希望中愉快，可惜的是，计划也总是赶不上变化，1965登场的“文化大革命”再次把束星北的希望抛进了深渊。

不过，对束星北来说，“文化大革命”虽然让他的“摘帽”希望再次落空，却也让他有幸脱离了焦点，毕竟，在这场更疯狂的运动中有更大、更刺激的鱼儿等待着革命群众去消遣。脱离了焦点的束星北渐渐无人问津，但他丝毫不敢怠慢，一丝不苟地执行着每天的改造任务。不过这时的他也开始偷偷读一些专业书，甚至趁扫雪的时候在雪地里写写公式。可是，脑子里的东西越多，无所事事的感觉就越折磨人。束星北是一个憎恶平庸，有着天才情结的人，随着年龄越来越大，眼看着自己的年华在遥遥无期的平庸中一点点耗尽，他心中的焦虑也日益加剧，终于忍不住再次向组织发出了请求，他写道：

我今年已64岁，改造了十几年还没有改造好。岁月蹉跎，心中焦急：如果再过十几年，即使改造好了，对党、对人民、对社会主义还能有什么用呢？……今后该怎么办，才能得到党和人民的宽恕、谅解和容纳？……恳请党领导、军工宣队能拉我一把，在我未死之前，……（让我）回到人民内部，尽自己的力量为巩固无产阶级专政服务……

他的请求虽然恳切，却终究是不可能缔造奇迹的。后来的发展显

示，他的真正出路，其实既不在于党和人民的“宽恕”，也不在于自己的“改造”，而在于外来的力量。这种力量终于来了：1972年10月，曾经是束星北学生的美籍华裔物理学家李政道踏上了久违的故土，成为继杨振宁之后又一位穿越“文革”铁幕的海外物理学家。李政道受到了高规格的接待，周恩来在会见他时请他为解决中国教育人才的断层问题做点工作。李政道表示，中国并不缺乏教育人才，而是没有给他们发挥才能的机会。作为例子，他说：“比如，我的老师束星北就在国内”。

李政道不仅提到了束星北，而且还表示要见一见这位启蒙老师。这样的动静吓坏了山东医学院的“革委会”，他们以右派分子不能进京的规定为由阻止了束星北上京。但问题是，束星北不去北京，李政道会不会来青岛，甚至提出要到束星北家去坐坐呢？要知道，束星北的家可不是一般的家，据一位“革委会”的成员回忆，束星北的家是这样的：

那是我所见到的最赤贫破旧的家，你说它家徒四壁吧，破破烂烂的东西似乎又不少：缺了腿的桌子（晚上便铺上被子做床用），两个箱子（部队装子弹的箱子），几张自己打制的歪歪斜斜的板凳和一些堆得乱七八糟的书；地板虽是水泥的，可是到处都是裂缝，客厅中间还有一个大洞，大得能陷下腿去，上面盖着一张三合板，简直就是个陷阱。最不堪的是束星北的“卧室”，他的卧室不过是个两三尺宽的壁橱，束星北的个头这么大，常年“卡”在里面能舒服吗？“卧室”里只有一床被子，严格地说，那不是被子，只是一床破破烂烂的棉絮，如不是一些经纬纬的黑色电工胶布粘连着，早就散了。

显然，这样的“教师之家”是万万不能让李政道看见的（万一李政道失足掉进那“陷阱”里，更是不堪设想）。为了解决这一难题，“革委会”绞尽了脑汁，考虑了N种方案，比如让束星北火速搬入原党委副书记的家里，让束星北住进宾馆等，都感到不保险，最后干脆快刀斩乱麻，以束星北身体不适为由推掉了此事。

李政道没能见到束星北，便给他写了一封信：

束先生：

自重庆一别，离今已有差不多廿八年了。对先生当年在永兴湘潭时的教导，历历在念。而我物理的基础，都是在浙大一年所建，此后的成就，归源都是受先生之益。

此次回国，未能一晤，深以为怅，望先生小心身体。

特此敬祝

工作顺利，身体健康

生李政道上

李政道虽未能见到束星北，但他的惦念还是给束星北带来了生机。收到李政道的来信（这封信束星北连拆都没敢拆，第一时间就上交给了组织）后不久，束星北又鼓起勇气给组织写了信，他在信中写道：

我自1957年向党猖狂进攻之后，经过十多年的劳动改造，思想有所改变，……因之悔恨交加，亟思以实际行动来取得党和人民的宽恕与谅解，……特上牒恳求给我立功赎罪

的机会……

一位如此弱小的学者，被迫承认向如此彪悍的党“猖狂进攻”，这实在很像是一则现代版的狼和小羊的寓言故事。幸好这党虽有一身横练功夫，却也有一个小小的练门，那就是特别器重外籍友人。束星北的待遇自李政道访华之后大为改善，他的“帽子”也终于在1974年被摘除，但他被“落实政策”则是在“文革”后的1979年，距离他被打成右派已有22年，距离他最早遭到批判则已隔了整整四分之一个世纪。



重返讲台的束星北



与老友王淦昌（右）在一起

“摘帽”后的束星北终于有机会重返讲台，但北大、厦大、暨大、中科院等依然不敢聘他，最后是国家海洋局第一海洋研究所打破禁忌聘请了他，成为他一生最后五年的归宿。残酷的政治运动让束星北历尽苦难，却并未真正改变他。“文革”后，很多科学家加入了党员的行列，但当海洋所的领导希望束星北也递交入党申请时，却遭到了断然的拒绝。复出后的束星北又见到了昔日的同事及好友王淦昌，两位垂暮的老人竟像年轻时那样为学术问题争得面红耳赤。临走时王淦昌感慨道：“你还是老样子。”束星北则回答说：“没办法，人是不可改变的。”

动荡的岁月终于过去了，可岁月留下的伤痕却再也无法抹去。晚年的束星北厌恶过去，避谈过去，却又时常陷入到过去的噩梦之中，难以自拔。他有时会忽然变得焦躁不安，梦游般地满屋子嚷嚷着寻找水桶、拖把和扫帚——那些都是他当年打扫茅房的工具。

束星北老了，在最后几年里，他拼命想把时间补回来，再完成些什么，那是他永不泯灭的天才情结所系，可惜一切已是镜花水月。1983年10月30日，这位际遇凄凉的前辈走到了生命之路的尽头。在去世前，他将自己的遗体捐给了青岛医学院，作解剖及制作骨骼标本之用。

束星北去世后，李政道、王淦昌、苏步青等著名学者发了唁电。束星北去世后的第三天，国家海洋局、青岛医学院及束星北的子女们举行了一个肃穆的遗体捐赠仪式。青岛医学院在遗体移交书上写道：我院对束星北教授这种献身精神表示敬佩。

但是，敬佩归敬佩，青岛医学院的很多人也正如束星北那样，是不可改变的。时势虽然变了，但每一次变迁都意味着风水可以轮转，权力可以重分，也意味着很多不可改变的人会投入新一轮的权力之争，因为那是他们永恒不变的兴趣。当然，每一次那样的争夺都师出有名，这一次的名目叫做“清除文革余孽”。等到硝烟散去、尘埃落定，终于又有人想起太平间里的束星北遗体时，已是半年之后的事了。人虽然不会变，遗体却是会变的，变得失去研究价值了。医学院的领导便遣两位学生将束星北的遗体送到一片荒凉的林子里去埋掉，而那两位学生嫌路太远，就近将遗体埋在了学校的篮球场边，那里如今立着一副双杠。

在那副双杠上锻炼的年轻学子们也许很少有人会知道，他们的双脚之下便是一位中国物理界前辈的埋骨之处。

二零零八年十一月六日写于纽约

(1) 束星北一生并未作出过有价值的学术工作，这与他后半生的悲惨遭遇、他的个性及他的研究方向均有一定关系（他早年曾致力于研究经典统一场论，那是一个铁定会失败的方向）。在人们争论束星北的学术成就时，曾有人举出束星北去世后李政道给束星北夫人的唁电中的话，“束老师是中国物理界的老前辈，国际闻名，桃李天下，他的去世是世界物理界及全国教育界极重大的损失”，来说明束星北具有很高的学术地位。但唁电（尤其是学生为自己启蒙老师所发的唁电）中的话是明显不能作为正式的学术评价的。

(2) 不过，作者对史料的广泛引用也有一个小小的缺陷，那就是对许多事件的记叙只有引文，而未标明时间。

(3) 除上述经历外，1979年3月9日的《光明日报》曾刊登宫苏艺整理，束星北署名的文章：《在爱因斯坦身边工作的日子里》。该文提到束星北曾于1928—1929年间在柏林大学（University of Berlin）爱因斯坦身边工作过一年。但此说后来引起了争议，从争议双方提供的信息来看，我倾向于支持怀疑方的看法，即认为此说不实。关于这一争论，以后有时间我将另文介绍。另外，束星北在游学过程中还曾到过日本、朝鲜和苏联等国（他对苏联的印象相当坏，认为那是一个专制和混乱的国家）。

泡利效应趣谈

按照最近一段时间我网站 (<http://www.changhai.org/>) 的更新频率，我决定在6月1日上传本文。不过“六一”这个日子似乎不太适合谈论严肃的话题，我们就挑一个以物理学家命名的非物理效应——泡利效应（Pauli effect）——来聊聊吧。

泡利（Wolfgang Pauli）是20世纪一位很早熟的理论物理学家（当然，他的早熟是学术上的早熟，跟“九零后”、“零零后”的早熟不是一个概念）。记得有人说过，理论物理与实验物理的鲜明区分是始于20世纪的，这种说法是否恰当有待考证，不过这种区分在出生于1900年的泡利身上倒是体现得很鲜明，且很有戏剧性。泡利的理论天赋无疑是很高的，但他的实验才能则不仅不高，而且有可能是负的。这种“负才能”的集中体现就是所谓的“泡利效应”——即泡利的出现对实验设备的破坏效应。



奥地利物理学家泡利

是谁最早采用了“泡利效应”这一名称？我不太清楚，不过这一效应在当时的实验物理圈中是颇有些名气的，而且那名气并非只是搞笑意义上的名气。泡利的朋友斯特恩（Otto Stern——1943年的诺贝尔物理学奖得主，Stern-Gerlach实验中的那位斯特恩）就曾因为担心泡利效应而对泡利下达了封杀令，禁止其进入自己位于德国汉堡的实验室。科学史作

品中有不少科学家逸闻是以讹传讹的故事（比如伽利略的斜塔实验、牛顿的苹果、瓦特的茶壶、爱因斯坦的板凳等），但这件是真的，斯特恩在1961年接受访谈时还亲自回忆过此事。

不过斯特恩的措施恐怕还小瞧了泡利效应的威力。据说有一次实验物理学家弗兰克（James Franck——1925年的诺贝尔物理学奖得主，Franck-Hertz实验中的那位弗兰克）位于哥廷根大学（University of Göttingen）的实验室里出现了一次事故。据查，泡利有不在现场的充分证据。于是弗兰克写信给泡利，很欣慰地告诉他说你总算无辜了一回。真的无辜吗？泡利对这种无罪推定似乎并不领情，他回信很诚实地“自首”说自己虽不在第一现场，但事发当时自己乘坐的从苏黎世到哥本哈根的火车却恰好在哥廷根的站台上停留了一会儿！——言下之意，泡利效应是可以有“超距作用”的。

泡利效应最初只是实验圈里的传说，但后来名头渐渐出了墙，波及的对象也变得五花八门，不再限于实验设备。据说有一次泡利去听一个讲座，他看见两位衣着体面的女士中间恰好有一个空位，便走过去坐了下来。谁知他刚一坐下，那两位可怜的女士所坐的椅子就双双垮掉了。这故事是德国裔美国艺术史学家潘诺夫斯基（Erwin Panofsky）讲述的，是真有其事还是艺术的虚构就不得而知了。

戈革先生也转述过一个类似的故事，说泡利在某次学术会议上听到了一个令他不满意的理论，便走上台去严厉批评。泡利的不留情面在物理学界是出了名的，他说到紧要之处时，忽然拿粉笔向那已经坐回台下的演讲人隔空指去，吓得那人向后退，结果那人所坐的椅子竟突然垮掉了。坐在那人背后的俄国物理学家伽莫夫（George Gamow）当即跳起来喊道：“泡利效应！”（后来有人根据伽莫夫“此地无银三百两”的那声吼，怀疑是他在报告人的椅子上做了手脚。）

有关泡利效应的传闻有一个有趣的特点，那就是泡利效应绝不会损害到泡利本人，不仅不会损害，甚至还能在关键时候“拯救”泡利。泡利的助手派尔斯（Rudolf Peierls）和韦斯科夫（Victor Weisskopf）都曾在自己的自传中讲述过一个故事（细节略有出入，但本质上是同一个故事），其情节大致是这样的：有一次泡利要参加一个学术会议，与会的年轻物理学家们决定跟他开个玩笑，他们在会议厅的门上做了一个触发式的机关，只要泡利一推门就会发出类似爆炸的响声。结果呢？在泡利推门而入的一刹那，那些被年轻物理学家们反复调试过的机关居然“卡壳”了！泡利效应通过破坏“实验装置”而成功地“拯救”了泡利。

泡利本人对泡利效应的这种“损人利己”的效果也很有信心，并将之作为判断一个效应是否为泡利效应的重要凭据。泡利后期的助手及传记作者恩兹（Charles Enz）曾经讲过一个故事。那是在1956年，泡利去世前的两年，恩兹夫妇与泡利夫妇在意大利的科摩湖畔共乘出租车时，车子在一个小山坡的顶上抛了锚。恩兹打趣说那是泡利效应在作祟。但泡利不同意，因为那个故障给泡利自己也造成了不便，不符合泡利效应损人利己的特点。

在所有与“泡利效应”这个金字招牌相联系的逸闻中，似乎只有一件是与泡利自己所受的伤害联系在一起的，不过那是一个伪效应（或者确切地说，是伪效应中的伪效应——如果考虑到泡利效应本身也是伪效应的话）。那件逸闻是这样的：泡利是个夜猫子，并且有一个不良嗜好，那就是喝酒（有时甚至是酗酒）。1932年夏季的某一天，他在美国安娜堡（Ann Arbor）参加暑期学校时喝醉了酒，从楼梯上摔下来，摔伤了肩膀。在接下来的一段时间里，他的讲座改由乌伦贝克（George Uhlenbeck——电子自旋概念的提出者之一）代写板书。据说那几次演讲由于不必亲自写板书，反而讲得特别精彩。不过泡利受伤的真正原因

是物理学家们的幽默而非真实的物理效应，这应该是相当显而易见的。但世界之大，却也不乏一本正经看待这种效应的人，比较著名的，我数了数共有两个半人（我很想拾日月神教任老教主的牙慧，找出“三个半”人来，可惜实在找不到能与那两个半人齐名的“第三者”）。这其中前文提到的斯特恩算半个——该同志虽相信泡利效应，但只知防范不求甚解，故只能算半个。剩下那两个“完整的人”（郭德纲语）一个是泡利本人，另一个则是瑞士精神科医生兼心理学家荣格（Carl Jung）。听起来有些令人难以置信，但像泡利那样出色的科学家在学术高峰期之后有时也不免像聂棋圣那样出昏招，而且那样的科学家还并不鲜见（因此太过器重——甚至举国器重——老科学家的暮年言论不见得是明智的）。泡利自30岁之后开始遇到一些心理方面的麻烦⁽¹⁾，在他父亲的提议下，他向荣格进行了咨询（他们的通信后来被整理出版）。自那个阶段起，部分地受到荣格的影响，泡利开始相信一些后来被人们视为伪科学的东西⁽²⁾。他不仅对泡利效应的真实性深信不疑，认为它可能与所谓的“超心理学”（parapsychology）有关，而且还写过文章，试图发掘物理学与精神分析学之间的“互补性”（complementarity）。泡利与荣格的交往对荣格也有一定影响，其信件被后者视为案例。荣格并且提出了诸如“非因果联系原理”（acausal connecting principle）、“非因果平行性”（acausal parallelism）之类的概念，试图解释包括泡利效应在内的一些“现象”。

尽管相信泡利效应的这两个半人都有赫赫的名头，不过若是让我来评论一下这几位爷在此事上的观点或做法的话，我愿意借着“六一”的气氛斗胆给出以下的评语：

- 斯特恩的做法：humourously wrong（错得幽默）；
- 泡利的观点：wrong（错）；

- 荣格的观点：not even wrong（连错误都不如）[\(3\)](#)。

当然，荣格的粉丝也别生气，就当是“六一”节的玩笑好了。

参考文献

[1] Enz C P. No Time to be Brief: A Scientific Biography of Wolfgang Pauli [M] . Oxford: Oxford University Press, 2002.

[2] Miller I. Deciphering the Cosmic Number: The Strange Friendship of Wolfgang Pauli and Carl Jung [M] . New York: W. W. Norton & Company, Inc., 2009.

[3] Peierls R. Bird of Passage: Recollections of a Physicist [M] . Princeton: Princeton University Press, 1985.

[4] Weisskopf V. The Joy of Insight: Passions of a Physicist [M] . New York: Basic Books, 1992.

[5] 戈革. 史情室文帚（下） [M] . 北京：中国工人出版社，1999.

二零零九年五月三十一日写于纽约

(1) 泡利在这方面的麻烦与生活中的不顺利事件有一定联系。1930年11月，他与结婚不到一年的妻子离了婚。令他愤愤不平（是真的不平，而非幽默）的是，使他妻子弃他而去的竟是一位普通的化学家，而不是斗牛士之类泡利无法匹敌的猛男（斗牛士的例子是泡利自己举的）。那真是“生可忍，熟不可忍”（韦小宝语），最低限度也是“伤自尊了”（宋丹丹语），而且这一事件似乎引发了他对自己在与女人相处方面的能力的担忧（有他的多封信件为证）。

(2) 不过泡利毕竟是泡利，哪怕在精神状态最低落的时期，他也依然作出了一些重要工作，比如撰写他的量子力学“新约”（Die allgemeinen Prinzipien der Wellenmechanik）。他后期在量子场论方面也有不俗的工作。

(3) “not even wrong”是泡利针对某位年轻物理学家的一篇论文发表的评语。后来有人将之引

申为与科学哲学上的可证伪性原理（**principle of falsifiability**）相联系的概念。从可证伪性原理的角度上讲，错误好歹算是被证伪的东西，“**not even wrong**”则是连可证伪性都不具有，因而连错误都不如。

让泡利敬重的三个半物理学家⁽¹⁾

好像我每次写泡利（Wolfgang Pauli）总是在“六一”附近，上次写“泡利效应趣谈”是在5月底，这次则是在6月初，可能是这个话题比较轻松吧。其实泡利的论文大都不是省油的灯，但他的故事却充满了轻松诙谐，在“六一”附近自然只适合谈论后者。

喜欢物理学史的朋友想必都知道，泡利是一位以批评尖刻、不留情面著称的物理学家，荷兰物理学家艾伦菲斯特（Paul Ehrenfest）曾给他取过一个外号，叫做“上帝的鞭子”（God's whip）。至于他的具体“执鞭”事迹，大家可能已经听得多了，在这里我只讲一个在我看来流传较少，并且特别有趣的。我们知道，俄国有位以狂傲著称的物理学家，叫作朗道（Lev Landau）。有人画过这样一幅漫画：朗道坐在讲台上，长着一对天使翅膀，头上绕着光环（仔细看的话，那光环似乎是用量子力学波函数 ψ 组成的），下面的学生则个个长着长长的驴耳朵，恭恭敬敬地聆听教诲。无独有偶，泡利的学生在泡利面前也有耳朵偏长的感觉，在泡利的学生中流传着这样一种说法，那就是他们可以问任何问题，而不必担心问题太愚蠢，因为任何问题对泡利来说都是愚蠢的⁽²⁾。



也许是看武侠小说种下的好奇心，当我第一次读到朗道的故事时，心中就闪过了一个问题，那就是：如果狂傲的朗道遇到尖刻的泡利，会

发生什么状况？

当然，我没指望这个准武侠问题会有答案。可没想到的是，这个问题居然真的有答案。据说有一次朗道到苏黎世去做演讲——众所周知，苏黎世是泡利的老巢，而且泡利当时正在巢内！结果一向狂傲的朗道藏起了天使翅膀（“隐形的翅膀”？），收起了 ψ 光环（波包收缩？），讲完后还破天荒地作绵羊状，谦称自己所讲的东西也许是错的。“噢，绝对不是，”泡利安慰说：“你讲的东西是那样地乱作一团，我们根本弄不清哪些是对的，哪些是错的。”

这个故事我在英文文献中没有看到过，也不知其原始出处，不过已故中国科学史学家戈革先生在他的《史情室文帚》中讲过这个故事，从而应该是有来源的——不过那来源也可能只是八卦。20世纪前30年的物理学史实在太过激动人心，距今虽然才不过一个世纪，却已像上古神话一样抹上了传奇色彩。一般来说，除非是物理学家在信件、访谈或回忆录中记述过的亲身经历的东​​西，别的一切都可能只是八卦。罗伯森

（Peter Robertson）在《玻尔研究所的早年岁月》一书中曾经写道：“在大多数情况下，这些逸闻几经转述，变得更具有寓言或象征性的含义，它们既是一代物理学家留给下一代的部分遗产，也是物理学史中围绕这一不寻常历史时期的不断增长的神话中的一部分。”

不过，事实也好，神话也罢，既然泡利是这样一位有趣人物，我们就要问了：这位以尖刻著称，甚至能让朗道变成“绵羊”的人物心中，是否也有让他敬重的物理学家呢？答案是肯定的，但人数不多。我数了数，总共只有三个半（上回在《泡利效应趣谈》中我以一人之差没能拾到任老教主的牙慧，这回不能再失手了）。不过在介绍这三个半人之前，我先要对“敬重”一词做一点说明，因为这个词对泡利和对别人是有一定差别的，对泡利来说，我觉得它的“定义”是这样的^③：

从不批评＝极度敬重
偶尔批评＝比较敬重
偶尔表扬＝有点敬重
狠狠批评＝正常朋友

好了，现在我们言归正传，先说那半个人，他的名字大家都很熟悉，叫做海森伯（Werner Heisenberg）。论辈分，他其实是泡利的师弟（年龄也比泡利略小）。尽管辈分不高，但早在1924年，在这位师弟尚未完成任何重大工作时，泡利就已对他刮目相看。当时人们正被复杂元素的光谱问题搞得焦头烂额，泡利在给玻尔的一封信中将几乎所有物理学家都损了一通，说他们可以分为两类，一类是先用半量子数算一遍，如果不行就改用整量子数；另一类是先用整量子数算一遍，如果不行就改用半量子数。但他特意加了一个注释：“我不把海森伯包括在内，他更有头脑。”泡利的遗孀在泡利去世后接受一位科学史学家采访时，也曾回忆说泡利对海森伯的物理直觉有很深的敬意，认为那种直觉“盖过了所有的反对理由”。



当然，泡利与海森伯也闹过别扭，在20世纪50年代，他们两人曾合作发展过一种非线性旋量理论（那是海森伯版的“统一场论”）。但泡利后来不仅退出了合作，而且对海森伯进行了公开且尖刻的批评。杨振宁曾回忆过1958年夏天他们在日内瓦国际高能物理会议上的争论，他说：“这是我从来没有见到过的，两个重要的物理学家当众这样不留情面地互相攻击。”但是，在同年秋天，泡利却告诉海森伯：“你必须把这项工作推进下去，你总是有正确的直觉。”当然，那段时间泡利的情绪和健康都已不太稳定（那年的12月15日，他就离开了人世），对他那段时间的言论也许不宜做太多解读。不过在与泡利年纪相仿的物理学家中，从未有第二个人从他那里得到过如此高的评价，因此海森伯这“半个人”的地位应该是很稳固的——或者说是“测得准”的。

现在来谈谈那三个“完整的人”（郭德纲语）。首先亮相的是“老牌劲旅”爱因斯坦（Albert Einstein）。虽然在有关泡利的故事中，不乏拿爱因斯坦垫背的，比如关于泡利指出爱因斯坦错误的故事，而且我还可以补充一件流传较少，但绝对真实的事情，那就是当爱因斯坦发表了反对哥本哈根诠释的EPR论文之后，泡利在给海森伯的信中曾经调侃道：“假如一个大学生在低年级时提出了这样的反对意见，我会认为他很有头脑，也很有前途。”⁽⁴⁾不过，尽



泡利与爱因斯坦

管留下了上述白纸黑字，综合地讲，泡利对爱因斯坦这位20世纪最伟大的物理学家是充满敬意的。我们知道，泡利和爱因斯坦晚年曾在美国普林斯顿高等研究院共事过。1945年，当泡利获得诺贝尔物理学奖时，普林斯顿的同事们为他举办了一个庆祝会。在会上，爱因斯坦出人意料地发表了简短的祝贺^⑤。泡利对来自爱因斯坦的这份祝贺极为珍视。几年后，爱因斯坦70岁生日时，泡利在给爱因斯坦的信中这样写道：

您的70岁生日给了我一个愉快的机会，在向您表示由衷祝贺的同时，告诉您我是多么感激您在普林斯顿给予我的私人友情，以及您1945年12月在研究院庆祝会上的讲话给我留下的记忆有多么难忘。

爱因斯坦去世后，泡利在给玻恩（Max Born）的信中，再次提到了爱因斯坦的那次讲话：

这样一位亲切的、父亲般的朋友从此不在了。我永远也不会忘记1945年当我获得诺贝尔奖之后，他在普林斯顿所作的有关我的讲话。那就像一位国王在退位时将我选为了如长子般的继承人。

泡利对爱因斯坦发自内心的敬重，在这些绝无半分调侃的文字中流露得非常清晰。

不过在让泡利敬重的人当中，爱因斯坦还排不到第一，甚至连第二

都悬，因为争夺第二的是一位强劲的对手：丹麦物理学家玻尔（Niels Bohr）。在量子理论发展史上，人们对玻尔具体贡献的大小或许有所争议（感兴趣的读者可参阅拙作《纪念戈革——兼论对应原理、互补原理及EPR等》的第三节和第五节），但几乎无可争议的是，他是一位极富感召力的领袖人物。大多数年轻物理学家对玻尔都非常敬重，其中包括那位以狂傲著称的朗道。朗道不仅敬重玻尔，而且还很谦虚地向他请教过有什么秘诀能把这么多有才华的年轻人聚集在自己周围（玻尔的回答是：我只是不怕在他们面前暴露自己的愚蠢。不过据说朗道向别人转述这一回答时，把“暴露自己的愚蠢”错说成了极具朗道特色的“暴露他们的愚蠢”）。作为哥本哈根学派的铁杆成员，泡利跟玻尔的关系也是非常密切的。1922年，玻尔在哥廷根做了一系列演讲，由此结识了海森伯和泡利。演讲后泡利在给玻尔的信中表示：“非常感谢您在哥廷根时那样亲切地让我了解到最广泛的问题，那对我来说有着无可估量的益处。”二十多年后，在回忆自己的科学生涯时，泡利再次表达了对玻尔的敬意，他说：“我科学生涯的一个新阶段始于我第一次遇见尼耳斯·玻尔。”



泡利与玻尔正在玩翻转陀螺（这是我最喜欢的物理学家相片之一）

关于泡利对玻尔的敬意，还可以补充一个很有趣的例子。我们知道，泡利曾把玻尔的弟子，比利时物理学家罗森菲尔德（Léon Rosenfeld）戏称为“教皇的唱诗童子”。但事实上，在玻尔这位“教皇”面前，泡利本人也时常表现得像一位试图博取疼爱的孩子。1924年底，在讨论元素光谱时，泡利给玻尔的信件中居然出现了这样的文字：“如果我的胡思乱想居然真能使您又亲自关心起多电子原子的问题来，那我就将是世界上最快乐的人了。”老实说，如果把上述文字中的物理去掉，我恐怕会分不清说话者是泡利还是琼瑶。世上能让泡利以如此语气写信的人，恐怕只有玻尔了。因此，玻尔在让泡利敬重的人当中绝对是可以坐三望二的（而且我倾向于把他排第二）。

不过，泡利和玻尔在多数观点上虽可算是同气连枝（与“站在了历史的错误一面”的爱因斯坦不同），但也免不了要受泡利的批评（当

然，大都是很客气的“治病救人”式的批评），比如玻尔那短命的BKS理论，就因为放弃了严格意义下的能量守恒定律而遭到了泡利的批评。后来当美国物理学家康普顿（Arthur Compton）等人用实验否决了BKS理论后，玻尔在给同事的信里承认，泡利“长期以来就是对我们的‘哥本哈根叛乱’不表同情的。”

那么，在与泡利有深交的物理学家之中，有没有谁是从未受过泡利批评——从而按我们的“定义”是让泡利“极度敬重”——的呢？有，但有且仅有一位，那就是泡利的导师索末菲（Arnold Sommerfeld）。事实上，泡利不仅从未批评过索末菲，甚至终其一生都在索末菲面前谨守着弟子礼仪。哪怕当他早已成为极有声望的物理学家之后，只要索末菲走进他的屋子，泡利就会立刻站起，甚至鞠躬行礼。他的这种乖顺的举止常常让习惯了严厉批评的他的弟子们“没事偷着乐”。对此，奥地利物理学家韦斯科夫（Victor Weisskopf）在其自传中有过很有趣的记述：

有一个人泡利对他的反应是不同的。当阿诺德·索末菲——泡利以前的导师——来到苏黎世访问时，一切就都变成了“是，枢密顾问先生，是，那是最有趣的，虽然我也许会倾向于稍稍不同的表述。我可不可以这样来表述？”对于太经常成为他霸气牺牲品的我们来说，看到这样一个规规矩矩、富有礼貌、恭恭敬敬的泡利是一件很爽的事情。

关于泡利对索末菲的敬重，除外人的观感外，还可以举出泡利本人的文字。索末菲70岁生日临近时，泡利开始撰写一篇索末菲感兴趣的文章。为了让文章能恰好完成于索末菲生日当天，泡利将完稿日期推延了几天。但擅自推延索末菲喜欢的文章是有风险的，万一老爷子生气怎么

办？于是他给索末菲写了一封安慰信：

您紧皱的眉头总是让我深感敬畏。自从1918年我第一次见到您以来，一个深藏的秘密无疑就是，为什么只有您能成功地让我感到敬畏。这个秘密毫无疑问是很多人都想从您那儿细细挖掘的，尤其是我后来的老板，包括玻尔先生。



泡利（右）与索末菲

事实上，不仅泡利如此，索末菲所有的学生在他面前都保持着恭敬的礼节，而索末菲本人似乎也很喜欢这种礼节。这种老派的德国礼仪曾让一些美国物理学家不以为然。比如美国氢弹之父泰勒（Edward Teller）就表示自己不太喜欢索末菲，原因就是太重礼仪。他说美国物

理学家冯·弗莱克（John Hasbrouck Van Vleck）年轻时曾经见过索末菲。第一次见到时，他打招呼说：“早上好，索末菲先生”，索末菲基本未予理会；第二次见到时，他打招呼说：“早上好，教授先生”，索末菲只是淡淡地笑了笑；直到后来当他说：“早上好，枢密顾问先生”时，索末菲才赞许地回答说：“你的德语进步很快啊。”⁽⁶⁾

但不管怎么说，在索末菲的学生之中，有六人获得过诺贝尔奖，几十人成为第一流的教授，他们的名字足可铺成一条20世纪物理学的星光大道。索末菲本人虽从未得过诺贝尔奖，却是一位无冕之王，是物理史上最伟大的教师之一，他在让泡利敬重的物理学家中拔得头筹是实至名归的。

也许在未来的某个“六一”，我会写写这位伟大教师的故事。

参考文献

[1] Enz C P. No Time to be Brief: A Scientific Biography of Wolfgang Pauli [M] . Oxford: Oxford University Press, 2002.

[2] Mehra J, Rechenberg H. The Historical Development of Quantum Theory [M] . Berlin: Springer, 2000.

[3] Seth S. Crafting the Quantum: Arnold Sommerfeld and the Practice of Theory, 1890—1926 [M] . Boston: The MIT Press, 2010.

[4] Weisskopf V. The Joy of Insight: Passions of a Physicist [M] . New York: Basic Books, 1992.

[5] 玻尔. 尼耳斯·玻尔集（第五卷） [M] . 戈革, 译. 北京: 科学出版社, 1991.

[6] 玻尔. 尼耳斯·玻尔集（第七卷） [M] . 戈革, 译. 北京: 科学出版社, 1998.

[7] 戈革. 史情室文帚（下） [M] . 北京: 中国工人出版社, 1999.

[8] 杨振宁. 杨振宁演讲集 [M] . 天津: 南开大学出版社, 1989.

二零一零年六月五日写于纽约

(1) 本文曾发表于《现代物理知识》2012年第4期（中国科学院高能物理研究所）。

(2) 绝对需要辨明的是，泡利和朗道的不留情面与吾国文人间的彼此相轻完全是两码事，前者充满了幽默和智慧，并且深得同行和学生的喜爱，而后者则是.....（此处作者省略137字）。

(3) 当然，上述“定义”只适用于泡利朋友圈中的人（包括师长、同事和学生）。跟泡利八杆子也打不着的人千万不要跳出来说：“泡利从来没有批评过我，因而他对我极度敬重”。

(4) 需要指出的是，泡利是特别喜欢在私信中使用调侃笔调的（尤其是调侃第三者——即收信对象以外的其他人），那对他是一种莫大的乐趣。泡利的前妻——就是我们在《泡利效应趣谈》的注释（[参考此处注\(1\)](#)）中提到过的那位嫁给化学家的前妻——曾回忆说，泡利“常常在我们房间里像一只笼中的狮子一样来回走动，构思着尽可能尖刻和诙谐的回信，这给了他巨大的愉悦”。一般来说，我对泡利私信的看法是：他私信中的赞许远比批评可靠，前者通常有确切理由，后者则往往只是调侃。

(5) 后来公布的文件显示，在泡利获奖前不久，爱因斯坦向诺贝尔奖委员会推荐了泡利。

(6) 索末菲的“枢密顾问”头衔究竟何时所获我始终未能查出。不过韦斯科夫的回忆录和泰勒的访谈都提到了此头衔，应该不是“克莱登大学”的。另有一种可能性是“枢密顾问”在德国可以作为敬称来用（就像中国的“大师”），不过我同样未能查到任何资料支持这种可能性。

纪念戈革——兼论对应原理、互补原理及**EPR**等

我的一生都没有休息过，不为别的，就为我喜欢的事情，不是为了稿费，不是为了别人，而是为了自己，我就是要翻译《波尔集》。

戈革（《正直者的困境》，2006年）

一、著作等身的翻译家

前不久，我从国内邮购了戈革先生的自选集《史情室文帚》。阅读之后颇有些感想及感慨，便打算写一篇文章，一来介绍一下这位造诣深厚却鲜为人知的玻尔研究者、科学史学家及翻译家（也许很少有人称呼他为“翻译家”，但我认为他当之无愧）；二来也评述一下他在玻尔的历史地位、对应原理、互补原理及EPR争论等方面的若干观点。结果在收集资料的过程中，却意外地得知老先生已于2007年12月29日离开了人世。



2001年美国物理协会刊登的戈革照片

过去这些天里，我几乎每天都在上下班的路上阅读戈革先生那些有个性、有见地、有火气，甚至有生命的文字，叹息于他的坎坷人生，钦佩于他的勤奋笔耕，为他对玻尔的眷恋之情而感动，也为他坦荡、率真及刚正不阿的性格而喝彩。可惜从今往后再不能看到他的新作了。就谨

以本文作为对他的纪念吧。

最早知道戈革是在1991年。那时我刚进复旦大学物理系，在图书馆里读到了他的著作《尼耳斯·玻尔：他的生平、学术和思想》。在当年的日记中，我写下了这样的感想：

这确实是一本不可多得的优秀作品。作为一位中国学者，研究近代物理学史的条件是很有限的。但戈革教授精心地对所掌握的材料进行分析、取舍，力求按照第一手资料，精确真实地再现玻尔生活的时代，这是这部书引人入胜的地方。在书中，随处可见作者对各种资料的评述及对自己历史观的阐述，这些阐述体现了作者独特的风格。在中国，历史研究常常因过多陷入政治的范畴，而导致单调、刻板、说教的风格。能够读到这样的一些富有逻辑、措辞严谨、尊重史实的文字，真有一种耳目一新的感觉。

后来我又陆续接触到了戈革的一些翻译作品，比如《原子理论和自然的描述》、《原子物理学和人类知识》、《原子物理学和人类知识论文续编》等。那些是他最早翻译的玻尔著作，他在20世纪50年代因看到苏联在批判玻尔及哥本哈根学派，便向商务印书馆自荐，表示愿意翻译玻尔作品，让大家知道其反动之处究竟何在，以便更好地进行批判。结果翻译之下却喜欢上了玻尔，而且翻译得越多，就喜欢得越多，最终一发不可收拾地将毕生精力都投入到了翻译与研究玻尔之中。戈革那几本早期译作出版于20世纪60—70年代。那时正值极左风暴横扫神州大地，喜欢玻尔是要被“砸烂狗头”的，因此不仅不能有丝毫流露，还要装模作样地与革命大众一起来批判玻尔。我当年在复旦看到的译作就是60—70

年代的旧版（当时戈革用了“郁韬”这一笔名），前面有批判玻尔的译者序。当初读到那些批判文字时我还觉得有点好笑，后来才渐渐体会到了译者所经历的那种人在江湖、身不由己的无奈与痛苦。1999年，那三本译作被合并再版为《尼耳斯·玻尔哲学文选》。戈革先生在再版序言中写道：

当年那三册小书的前两册出版于“十年动乱”以前，而其第三册则属于我所谓的“牛棚译”，那是在九死一生、众叛亲离、朝不保夕、饱受折磨的非人生活中偷偷译成的。

“九死一生、众叛亲离、朝不保夕、饱受折磨”——这短短十六字的背后埋藏了多少血泪？正所谓“多少辛酸事，尽付十六言”。戈革将当年这种偷偷译书的努力称为“镇痛剂”。我想，在那暗无天日的岁月里，也许正是玻尔伟大的人格力量，哥本哈根令人神往的学术氛围，为在污浊人世中苦苦挣扎的戈革保留了心中仅有的暖意。或许也就是在那时，玻尔研究成为了他一生的心灵归宿。

1988年秋天，戈革应尼耳斯·玻尔文献馆（Niels Bohr Archive）馆长伊瑞克·吕丁格（Erik Rüdinger）的邀请，到丹麦访问了半年。那是他首次访问丹麦。在那里，他两度前往阿塞斯腾教堂公墓（Assistens Kirkegaard）拜谒玻尔墓。回国后，他在《自然杂志》上发表了八篇短文，记叙自己在哥本哈根的见闻。其中对第二次拜谒玻尔墓，他留下了这样的记叙：

阿塞斯腾教堂公墓中葬有丹麦近代的三大文化名人，即

诗人、童话作家安徒生，哲学家、神学家基尔凯郭尔和物理学家、哲学家玻尔。在我即将离开哥本哈根之际，已经成了我的好朋友的伊瑞克说，我们应该再去一次阿塞斯腾公墓，以便看看另外两位名人之墓。我欣然同意。这一次（1989年2月19日），我买了一盆鲜花，先到玻尔墓前献花告别，总算得偿了第一次谒墓时未能完成的献花心愿。但因在玻尔墓前流连得太久，墓园即将关门（下午5时），所以到底没能到另外两位名人墓前一拜，真所谓缘慳一面了！

这段话寥寥数语，细品起来却是真情流露，带着一种拜谒亲人般的依恋。而这种依恋，又浸透着无比的孤单。他在《史情室文帚》扉页上所印的就是一张摄于玻尔墓前的相片——在冬日的萧瑟树林中，老先生一人孤零零地站立在墓碑旁。

戈革先生是中国最杰出的科学文献翻译者之一，一生所译著作超过1000万字，其中以《尼耳斯·玻尔集》（Niels Bohr Collected Works）最为著名。这套总计十二卷的文献他以一人之力译出了十卷，而这十卷译作的出版更是历经磨难，几度遭遇出版社的出尔反尔，也几经更换出版社。1990年4月，戈革曾写过一篇短文《十年辛苦自家知》，叙述自己为翻译和出版《尼耳斯·玻尔集》十年劳碌奔走的甘苦。那篇短文的发表过程本身似也在为其所承载的辛苦



1988年戈革摄于哥本哈根玻尔墓旁

主题做着诠释。戈革在将该文收入《史情室文帚》时写道：

此文脱稿于1990年4月4日，刊于《博览群书》的某期，是该刊一位编辑向我约的稿。刊出时被编辑部粗暴删改，将后面的很长一部分完全砍掉。结果前面说《玻尔集》有“三方面的”意义，而后面却只论述了“两方面”（第三方面被删掉了），呈现了“前言不搭后语”之势。这样的“删节”实在开了我一个“大玩笑”！

“十年辛苦”还只是辛苦的开头，随着戈革翻译生涯的继续，出版的艰辛也在继续。首先遇到的当然就是钱的问题。早在1980年，戈革就颇有先见之明地在一篇发言稿中写道：

“经费”从哪里来？有没有“利润”？这样的问题是我这种书呆子所最害怕和最外行的东西。我只希望，如果咱们确实真心实意地搞“现代化”，钱（！）的问题应该有办法解决。不然呢？——那就一切全都等于废话！

现在看来，“咱们”搞现代化倒是真心实意的，但“咱们”也真心实意地不想把钱扔在这些无法生出更多钱来的地方。最终，还是靠他这位“书呆子”从国外筹来了经费，将《尼耳斯·玻尔集》的出版惨淡延续了下去。但即便如此，命运依然要残酷地让他在有生之年见到这套自己为之付出无数心血的文集被排挤出中国大陆。等他完成第十卷的译稿时，偌大个中国大陆已找不到一家合适的出版社。该卷最终由他自费，

并采用香港书号出版，由此导致的直接后果，便是无法在内地销售（因为没有中国大陆的书号）。该卷的很多存书也因此而一直堆积在戈革家中。

戈革翻译《尼耳斯·玻尔集》的努力在中国遭遇了重重困难，在玻尔的故乡丹麦却引起了很大反响，并受到媒体的广泛报道。戈革生前曾三度应邀访问丹麦，每次历时约半年。除学术交流外，他还利用国内外赞助的经费收集整理了大量与玻尔有关的文字、图片、影像及实物资料，这些资料起初大都收藏于戈革生前所在的中国石油大学的玻尔文献室里，戈革去世后，该文献室变得越来越难以维持，最终于2010年4月迁往了清华大学。

2001年夏天，戈革获得了丹麦国旗骑士勋章（Danish Order of the Dannebrog），成为继《安徒生童话》的译者之后又一位获此殊荣的中国人。那年秋天，美国物理联合会（American Institute of Physics）在一篇通讯中报道了戈革，并称其为（研究）科学史的英雄（A Hero of History of Science）。在戈革去世前不久，他收到了《尼耳斯·玻尔集》的最后两卷，那时他因眼疾加剧，已无法独立从事翻译工作，但依然凭借放大镜及自费聘请的助手的帮助，顽强地完成了译文的初稿，使这套第一流的译作不至于成为断臂的维纳斯。⁽¹⁾

除《尼耳斯·玻尔集》外，戈革还翻译了玻尔、海森伯（Werner Heisenberg）、爱因斯坦（Albert Einstein）等人的几种传记及其他相关文献。他也翻译过《爱因斯坦全集》（*Collected Papers of Albert Einstein*）的第三卷、麦克斯韦（James Clerk Maxwell）的《电磁通论》（*Treatise on Electricity and Magnetism*）等名著。他不仅英文功底深厚，而且翻译态度严谨，常常会为搞清一个典故而去信向国外友人询问。更重要的是，他的翻译大都是在读懂原著的基础上进行的，这可以

从他在译者序或译者注中对原著所作的精辟评论中看出。他对玻尔著作的研读更是细致入微，在去世前口述的自传体文章《正直者的困境》中，他对自己的翻译曾有过这样的叙述：

说到《玻尔集》，也有人劝我把剩下的工作交给别人翻译，但我不放心，因为那不仅是英语，而且是丹麦的英语，是玻尔的英语。我翻译了十卷，我比较了解他运用英语的方式，英语在他身上，用法不太一样。

毫无疑问，译作是戈革先生留给中国史学界最宝贵的遗产。

二、经历坎坷的独行侠

现在再回过头来叙述一下戈革先生的生平。戈革先生的一生，最主要的工作是研究科学史，尤其是研究玻尔。但正如他自己所说：“我研究玻尔，逾三十年，但在国内受到封锁和排挤，可谓全无任何知名度。^②”与所有“全无任何知名度”的人一样，有关戈革的生平，可资参考的材料极少。他所任教的中国石油大学数理系曾有一段简介，提到他1922年1月22日出生于河北省献县农村（其他资料一般只提到他出生于1922年1月）。他一生著作等身，却极少述及自己的生平。晚年时曾有朋友劝他写一本自传，但他却想等玻尔的书翻译完了再写。可惜上苍并未给他那么多的时间。现在我所能找到的有关戈革生平最有价值的材料，是他2006年口述的自传体文章《正直者的困境》，其余信息则只能从散见于其他文章的只言片语中进行收集与推测。

戈革幼年丧父，童年岁月在颠沛流离中度过。15岁那年，抗日战争爆发，戈革在乡下躲避了两三年，此后辗转于河南、甘肃等地，在甘肃酒泉念了中学。1945年，戈革考入了当时中国最好的大学——或许也是中国教育史上最令人景仰的大学——国立西南联合大学（简称西南联大）。一年后，西南联大迁回北大、清华及南开旧址复校。戈革选择进入了北大。

1949年，戈革从北大毕业，进入清华大学当研究生，但因当时教学秩序混乱，离校时并未获得学位。毕业后，虽经努力，志在理论物理的戈革仍未能进入综合性大学从事理论物理的教学及研究，而是被分配到了工科学院——先是进了山东工学院，后又于1953年进了当时刚刚成立的北京石油学院（现已改名为中国石油大学），在那里工作了一辈子。我最早知道戈革的时候，对他的石油学院教师这一身份颇感意外，因为

这个学院与他的专业、兴趣及著作等简直八杆子也打不着。他的这一身份后来还让人闹出过笑话。在戈革接受丹麦国旗骑士勋章时，曾有位中方领导到场发言，此人并不认识戈革，听说他是石油大学的教师，就在仪式上大谈了一通丹麦和中国的石油问题，并赞扬了戈革对中国石油业的贡献。

像戈革这样理论出身，在工科院校“不务正业”、“理论脱离实际”的教师，在政治运动中自然难逃革命学生、革命同事及革命领导们的“法眼”，先是众望所归地被划为右派，后又于1969年被定为反革命，在牛棚中度过了漫长岁月。在那些受尽屈辱，并且丧失行动自由的黑暗日子里，他的母亲离开了人世⁽³⁾。

戈革的一生历经坎坷，但那样的经历并未将他打磨成一个胆小懦弱、圆滑世故的人，相反，他脾性顽强，性格火爆，言辞犀利，嫉恶如仇。他的文字火气十足，火力更是十足，但同时却又不乏幽默，富有机锋。他对学术界外行领导内行、不学无术、信口开河、向壁虚造等种种弊端及腐败深恶痛绝，在文章中常予以毫不留情的痛斥。他在1980年的“科学技术史学术会议”上发言提出：

多年来对我们学术风气污染最严重的，就是“江湖气”和“行帮气”。这两者都带有很浓烈的封建味道，而且往往形影不离，交相为用。学术界的一切以假乱真、装腔作势、夸夸其谈、洋洋自得、投机取巧、妒贤嫉能、剽窃抄袭、结党营私、招兵买马、独霸一方、开山立柜……种种的恶习，几乎都不外是这“二气之‘良’能”！

从某种意义上讲，戈革是“文革”之后公开批评学术腐败的先驱者之一，这种批评在当时是需要很大勇气的。由于他这种火爆的脾气和嫉恶如仇的性格，戈革在周围很多人眼里成为了一位怪人，日益受到孤立和排挤。用他自己的话说：“我是一个很不好惹的老家伙……一辈子批评别人不看头势，常被人家恨得‘牙痒痒地’。”他在国内的知己寥寥无几，远少于他在海外的朋友。

除泛泛抨击学术界的腐败、肤浅、跟风、出丑等行为外，戈革也通过自己的科学史研究，具体指出并纠正了许多长期存在的以讹传讹、广为流传（或有意散布）的错误。比如他曾撰文驳斥所谓玻尔的互补原理与中国古代的阴阳符号或太极八卦存在关联的无稽之谈。他也曾驳斥过那些认为《易经》或《博物志》等中国古籍包含现代科学思想的“研究”。

当然，戈革关注得最多，并且研究得最透彻的，还是与玻尔有关的科学史。在这方面，他所指出并纠正的某些错误不仅普遍存在于中国史学界，也存在于国际学术界。比如有关玻尔当年面对的原子稳定性问题的真正含义，绝大多数中外文献都只给予了片面描述；比如许多中外教科书在提到玻尔原子模型时都有所谓玻尔的“三条假设”之说，这与玻尔原始论文的结构明显不符；又比如对玻尔对应原理的理解，绝大多数中外文献都存在着明显的曲解。[\(4\)](#)



除此之外，戈革对物理学史（尤其是量子力学发展史）的其他方面（比如玻尔是否如一些史学家认为的那样反对过爱因斯坦的光量子假说），以及一些疑案（比如发生在1941年的玻尔、海森伯密谈事件的真相），也都有所涉猎，并翻译过一些文献。值得一提的是，戈革为自己译作所写的译者序是他文字中很重要的组成部分。他的很多译者序不仅是精彩的书评，更是自己的研究心得，读起来引人入胜，而且很有参考价值，是阅读戈革译作独有的一种额外乐趣（只可惜有些译者序因言辞过于激烈而惨遭出版社编辑的毒手）。

戈革在翻译及研究玻尔方面的成果，赢得了国外玻尔研究者们的尊

敬，在彼此的交流中他与其中的很多人成为了好友。玻尔唯一从事物理研究的儿子，1975年诺贝尔物理学奖得主奥格·玻尔（Aage Bohr）甚至认为戈革也许能当互补原理的“亚圣”，即除玻尔本人外最深知互补原理含义的人。当然，戈革认为那是奥格·玻尔跟他开的“丹麦式的玩笑”，但即便是玩笑，能让奥格·玻尔开这样玩笑的人恐怕也不多。戈革的学术活动得到了海内外一些机构及个人的资助，他将这些资助都用于了学术活动及资料收集，他自己的生活则一直很清贫。他81岁时曾在一篇文章中写道：

老夫一生最向往的，便是一间比较宽敞适用的书房，但至今未得，而且今后也不会有任何希望……多年以来……虽然也有自己的工作场地，但是那都称不得书房，最多只是一间斗室，兼作读、写、会客、储藏（衣服、被褥）和养猫之用而已。此种场地有其自己的变迁，地址、结构和大小各不相同，最小之时乃是床头约两平方米之地，只放得下一桌一椅而已（我所译《玻尔集》的头四卷便是在该处完成的）。

戈革先生的这种经历，是不少老一辈知识分子（尤其是那些不会趋炎附势的知识分子）的共同经历，也是中国亏待那一代知识分子的明证。

除物理学史外，戈革先生在古体诗、篆刻、书法、绘画等方面也有出众的才能，限于本人的知识范围，本文对这些就只能一概从略了。在本文接下来的部分中，我将对戈革先生有关玻尔的历史地位、对应原理（correspondence principle）互补原理（complementarity principle）及EPR争论等方面的若干观点进行评述，并提出一些不同看法。戈革先生

若泉下有知，当不会介意我在一篇纪念他的文章中提出一些不同看法，因为国内常见的那种如样板戏一般高、大、全的浮夸传记或自传也正是他本人所素来不喜的。

三、关于玻尔的历史地位

在戈革先生的观点中，最有争议的也许是对玻尔的历史地位，尤其是对他与爱因斯坦的相对地位的评价。戈革在这方面的基本观点，是认为玻尔的历史地位不在爱因斯坦之下。^⑤比如他在《尼耳斯·玻尔：他的生平、学术和思想》一书的序言中这样写道：

真正全面地、综合地、历史地、不偏不倚地考察起来，无论就其深度、广度、强度来看还是就其持久性来看，玻尔在20世纪物理学中乃至在整个的人类思想领域中所起的作用，无论如何不亚于爱因斯坦。

又比如他在《尼耳斯·玻尔的学术贡献》一文中，对玻尔的评价是：

从他们对物理学本身的全部贡献来看，玻尔无论如何不亚于爱因斯坦。从他们在20世纪物理学革命中所起的推动作用来看，玻尔显然超过爱因斯坦。

这两个“无论如何不亚于”和一个“显然超过”很清楚地表明了戈革对玻尔历史地位的评价。而在被他视为“学术遗嘱”的《〈尼耳斯·玻尔集〉译后记》中，他更是毫不含糊地提出：“尼耳斯·玻尔在科学史上的地位，可以和牛顿的地位相比。”

戈革对玻尔的推崇遭到了很多人的抨击。他60岁时（1982年）甚至在一次演讲中因这一观点而被一位比他更老的“老先生”轰下台去。后来他在文章中愤然评论道：

在国外，爱因斯坦的名声被无知大众吹得甚高，这未必是爱因斯坦之福（有关议论在此无暇分析）；在国内，个别人不许任何人对爱因斯坦有任何批评，谁要说了一句不敬的话，他就会暴跳如雷，那一副学阀面貌也很好玩！

此外，有批评者认为戈革对玻尔的推崇，是典型的研究什么就抬高什么。更有人指责他借抬高被研究对象来抬高自己。在我看来，戈革对玻尔的推崇是发自内心的，是因认为其伟大而研究，而非相反，更不是为了抬高自己。他在各种细节上流露出的对玻尔的感情，不是一个怀着功利目的而从事研究的人所能具有的。至于那种将戈革轰下台去的粗野做法，则不值一评。

不过，撇开各种情绪化的因素，我们也应该看到，爱因斯坦的名声绝非只是“无知大众”的吹捧，事实上，就连戈革本人也在某些文字中承认爱因斯坦是“实至名归”。在我看来，如果一定要对科学家的历史地位进行排序，那么20世纪的物理学家中唯一没有悬念的就是第一名，即爱因斯坦（类似地，20世纪之前的物理学家中唯一没有悬念的也是第一名，即牛顿），这或许也是多数人的共同看法。那么，戈革究竟是出于什么理由，认为玻尔的历史地位“无论如何不亚于”，甚至“显然超过”爱因斯坦呢？从我收集的资料来看，他的依据主要有以下四条：

（1）因为量子理论比相对论更重要。

(2) 因为量子理论对观念的突破比相对论更深刻。

(3) 因为玻尔对同时代物理学家的影响更大。

(4) 其他零星理由。

在本节中，我将对上述四条（尤其是二、三两条）作一些分析。鉴于此类问题固有的模糊性，这些分析当然不可能逻辑地推翻任何观点，因此本节的侧重点将是对上述几条的内涵进行分析，并提出一些在我看来是利用上述几条进行分析时不可忽视，却往往被忽视的地方。考虑到上述几条（主要是前三条）在文献中常常出现（虽然用意各不相同），对它们的内涵进行分析或许要比得到或推翻一个结论更有意义。至于结论——即那个见仁见智的排序问题——就像《雪山飞狐》中胡斐那一刀是劈还是不劈，每位读者可以自己品味。

上述四条中的第一条人们已讨论得很多，我不拟多说。总体来讲，量子理论比相对论更重要，是几乎所有物理学家的共识。不过，用这一条分析玻尔与爱因斯坦的相对地位时，最大的困难在于：相对论的创立基本上是爱因斯坦的“独角戏”（即便狭义相对论，他也是基本独立——甚至孤立——地完成的），而量子理论的创立却是一大批物理学家——其中包括爱因斯坦——共同努力的结果。因此从量子理论与相对论的对比到玻尔与爱因斯坦的对比，其间存在巨大的鸿沟。

接着看第二条，量子理论对观念的突破比相对论更深刻，这也是几乎所有物理学家的共识。用这一条分析玻尔与爱因斯坦的相对地位时，除面临与前一条同样的鸿沟外，在我看来还有一个不可忽视，却往往被忽视的地方，那就是在量子理论的发展中，观念的突破往往滞后于理论框架及实验结果的出现。比如普朗克（Max Planck）是首先通过含义

不清的数学内插手段得到与实验相符的黑体辐射公式，然后才提出量子假说；比如玻尔在构筑原子模型时，关注的是直接来自观测的原子光谱问题而非诸如“当电子从一个定态过渡到另一个定态时，它怎么决定将以什么频率来振动”之类的观念性问题；比如导致巨大观念性困难的波粒二象性，主要是由实验确立的；比如颠覆了经典决定论的波函数几率诠释，主要来自散射实验方面的经验。⁽⁶⁾与之不同的是，在相对论的发展中，观念的突破却先行或同步于理论框架，且独立于实验结果。比如爱因斯坦在建立狭义相对论的过程中，一开始就对诸如同时的相对性之类的观念性问题给予了极大的重视。⁽⁷⁾

因此，量子理论对观念的突破虽然比相对论更深刻，但它的发展过程却是由一系列比相对论的发展过程更接近常规的步骤组成的，相应的观念性问题则往往是在那些接近常规的步骤完成之后才出现或引起重视的。换句话说，量子理论的观念性突破往往是在以接近常规的方式打开理论框架及实验结果这个潘多拉盒子之后，才自然而然跑出来的（这时候想赶也赶不走了），是一种“逼上梁山”的被动式突破。虽然思考、分析及接受这种突破同样需要巨大的勇气与智慧，而玻尔在其中起了重要作用。但它与爱因斯坦创立相对论所经历的主动式突破在难度、创造性、贡献度等方面该如何对比，是一个需要考虑的因素。

再看第三条。这一条本身显然也是成立的。玻尔与同时代的很多物理学家——尤其是年轻人——有着密切的讨论关系，在这点上他绝非本质上是“个体户”的爱因斯坦可比。⁽⁸⁾可以说，玻尔是量子理论发展史上的领袖人物。不过，由讨论所产生的影响是一个高度含糊的概念，量子理论后来的发展在多大程度上得益于玻尔的这种影响，是一个很难确切回答的问题。退一步讲，即便玻尔能因这种影响而获得明确的功劳，也必须看到，由德布罗意（de Broglie）的物质波、薛定谔（Erwin

Schrödinger) 的波动方程以及玻恩 (Max Born) 的几率诠释组成的波动力学的主线是独立于玻尔发展起来的 (这条主线倒是在每个环节上都受到爱因斯坦的某种影响)。同样地, 狄拉克 (Paul Dirac) 的相对论量子力学也是独立于玻尔发展起来的。不仅如此, 就连量子力学的另一条主线——矩阵力学——的提出与玻尔也没有显著关系, 虽然矩阵力学的创始人海森伯 (Werner Heisenberg) 与玻尔有着密切关系。《尼耳斯·玻尔集》第五卷的编辑吕丁格 (Erik Rüdinger) 和斯陶耳岑堡 (Klaus Stolzenburg) 在有关矩阵力学及电子自旋的部分 (即该卷第二编) 的引言中这样写道: “一个很奇特的事实就是, 尽管我们在这儿讨论的这个时期在其他方面是以玻尔和海森伯之间的密切相互作用为其特征的, 但在1925年夏天却很引人注目地没有任何这样的作用。尽管海森伯在给克朗尼希 (Ralph Kronig) 的信中尤其是在给泡利的信中详细叙述了他的进展, 但是现存的信件中却几乎没有包含关于玻尔如何得悉矩阵力学的证据。”

另一方面, 玻尔是一个讨论型的物理学家, 讨论对于他自己中期及后期的工作也是至关重要的, 因此这种因讨论而产生的影响具有双向性。如果要谈论功劳, 其实也必须考虑这种双向性。

除此之外, 我觉得在用这一条分析玻尔与爱因斯坦的相对地位时还有一个不可忽视, 却往往被忽视的地方, 那就是玻尔与年轻人的讨论有时也会产生负面影响。比如荷兰裔美国物理学家兼科学史学家派斯

(Abraham Pais) 在《尼耳斯·玻尔的时代》(*Niels Bohr's Times*) 一书中就提到, 与像玻尔那样不知疲倦的人合作时, 年轻人几乎没有任何时间和精力来搞自己的研究。^⑨他并且引述合作者之一的瑞典物理学家克莱因 (Oskar Klein) 的话说, 他是当玻尔不在哥本哈根时才做出自己最独创和最大胆的工作的。无独有偶, 海森伯也是在和玻尔讨论得筋疲力

尽之后，趁其不在哥本哈根时完成了有关不确定原理的论文的，对这一工作有过启示的人是泡利（Wolfgang Pauli）、玻恩、约尔当（Pascual Jordan）及爱因斯坦，但不是玻尔。举这些例子，当然不是要抹杀玻尔对年轻人的正面影响，而是要强调在评估这种影响时所需面对的复杂性。

关于第四条，我们在这里只讨论其中特别有趣的一个观点。⁽¹⁰⁾戈革在其为《科学家传记词典》撰写的词条中曾经写道：

甚至有人说，玻尔早就有了关于新量子力学的概念，只是他那种无休止地修改文稿的办法，才使他延误了时间，落在了别人的后面。

这段引文似在暗示玻尔有能力自己提出新量子力学。但事实上，玻尔在博士毕业之后的几乎所有工作中，都明显呈现出定性分析越来越多，定量计算越来越少的趋势。⁽¹¹⁾而无论矩阵力学还是波动力学，它们最大的特点是奠定了量子理论的数学框架。如果海森伯和薛定谔没有作出他们的发现，或许早晚会有其他人脱颖而出，但这“其他人”却几乎不可能会是玻尔。因为玻尔虽对当时的整个形势有广博而深入的了解，但以他那种偏向定性的研究风格，是极不可能得到像矩阵力学或波动力学那样具体的数学框架的。因此，上述引文有很大的水分。

除上述疑似“水货”的引文外，戈革还在讲述新量子力学的诞生时，援引了玻尔最忠实的学生，被泡利称为教皇（指玻尔）的唱诗童子的比利时物理学家罗森菲尔德（Léon Rosenfeld）的一段话：

玻尔观点的充分影响一直停留在他的为数很少但却很有才华的一群学生中间，这些弟子们在通过及时地发表自己的结果而使自己更广泛地为人所知方面确实比他们的老师干得更好。

戈革并且评论道：“这句颇有‘微言大义’味道的话确实是很值得我们玩味的。”引用这样的“微言”，其“值得玩味”的“大义”似乎也是要暗示玻尔本人若不是发表文章的速度太慢，是有可能亲自创立新量子力学的。而这——如我们刚才所分析的——恐怕只能算是“唱诗童子”所唱的圣歌，是对玻尔极大的高估。

四、关于对应原理

对应原理（也叫对应论证）是旧量子论时期玻尔的一个重要贡献。在矩阵力学出现后，玻尔曾表示：“量子力学的整个工具，可以看成是对包含在对应原理中的那些倾向的一种精确表述。”玻恩也认为，对应原理是从经典力学通向量子力学的桥梁。但对于究竟什么是对应原理，文献中却存在着广泛误解。这一点，戈革先生曾反复指出过。

文献中对对应原理最常见的误解有这样几种：

（1）对应原理是指：量子数很大时，量子理论应过渡为经典理论。

（2）对应原理是指：普朗克常数趋于零时，量子理论应过渡为经典理论。

（3）对应原理是指：任何理论，在应用于经典理论可以适用的情形时，必须过渡为经典理论。

这些误解流传极广，几乎所有学过量子力学的人对上述表述都会有似曾相识的感觉。戈革曾在《尼耳斯·玻尔和他的对应原理》一文中列举了十类包含误解的文献，这其中包括了像派斯那样的物理学家撰写的文章。在我的藏书中也可以找到很多类似的误解，并且也都来自物理学家。比如美国物理学家玻姆（David Bohm）在其《量子理论》（*Quantum Theory*）一书中这样叙述对应原理：量子定律必须这样来选择，即在涉及众多量子的经典极限下，量子定律的平均结果应导致经典方程。又比如法国物理学家德斯帕那特（Bernard d'Espagnat）在《量子力学的概念基础》（*Conceptual Foundations of Quantum Mechanics*）一

书中认为，对应原理指的是当量子不连续性可被视为无穷小时，量子理论与经典理论对可观测量的预言应当重合。这些说法大致对应于上述第一、二两种误解。德斯帕那特并且把他的同胞、法国物理学家梅西亚（Albert Messiah）也拉下了水，他说后者的经典教材《量子力学》（*Quantum Mechanics*）包含对应原理的一种“稍微不同的想法”，而那个“想法”其实乃是上述第一种误解。

一般来说，戈革在指出他人错误时所用的口气是很不客气的，虽不能与如今互联网上的笔战风格相比，却也颇为尖锐。这一点在评述有关对应原理的误解时也不例外。不过值得一提的是，戈革在批评误解对应原理的人时，对误解者作了区分，将物理学家单独列为一类。他认为物理学家的兴趣不在于回顾过去，在历史问题上的误解最可同情和原谅。能作出这种区分，是难能可贵的，因为很多搞冷门研究的人，喜欢拿一些自己擅长的犄角旮旯的东西来抨击别人，通过别人（尤其是名人）在少数冷门内容上的疏失，来对他们进行语气夸张的否定，却全然不顾那些东西在别人的论述中是细枝末节还是核心主题。而戈革既指出了问题，又不夸大问题，意识到同样的问题对不同的人或在不同的作品中，其重要性是不同的，有些——比如随口提到对应原理的量子力学教材——是出现在蛇足上，有些——比如专论对应原理的科学哲学或科学史作品——则是出现在七寸上。他的批评更多的是针对后者，这是一种比较清醒的史学视角。

那么究竟什么是对应原理呢？戈革给出的表述是这样的：

对应原理是指：对于从定态 n' 到 n 的跃迁：

（1）当 n 与 n' 都是大数，且都比 $n'-n$ 大得多时，跃迁辐射的频率与相应轨道的经典绕转频率的某个倍频（或者说与一个经典泛频）基本重

合。特别是，当 n' 与 n 为相邻定态时，跃迁辐射的频率基本等于相应轨道的经典绕转频率。

(2) 在上述大量子数极限情形下，跃迁几率正比于相应泛频所对应的经典振幅的平方。

(3) 上述规律不仅适用于单周期体系，并且也适用于多周期体系。

(4) 上述规律不仅在大量子数极限下成立，而且或许在任意量子数下也成立。

与之相比，包括前面所提文献在内的多数文献的表述至多只包含了上述含义的前三条，有的甚至只包含了第一条——这一条在玻尔1913年的原始论文中曾被用到过，这可能是那些文献将之误解为对应原理的原因。但玻尔从未将上述第一条或前三条单独称为对应原理，相反，他曾对罗森菲尔德表示：“那不是对应论证。对于低频振模来说量子理论应该变成经典描述，这样的要求根本不是一条原理。它只是对理论的一种显然的要求。”因此，对应原理的真正关键实际上是在被多数文献所忽略的第四条上。正是因为有这一条，玻尔才可以表示“量子力学的整个工具，可以看成是对包含在对应原理中的那些倾向的一种精确表述”。否则的话，对应原理变成了只针对大量子数极限，它的“精确表述”又岂能成为“量子力学的整个工具”？

但是，戈革对上述第四条的表述却并不正确。因为如果那个表述正确的话，整个原子光谱理论除定态条件外，就完全变成经典物理了，这显然是不可能的。因此，由对应原理的前三条所表述的规律，事实上还真的“仅在大量子数极限下成立”，在任意量子数下是不成立的。这一点

玻尔早在1918年首次正式表述对应原理时就已意识到了。[\(12\)](#)因此戈革对这一条的表述（这里所选的已是戈革文章中最谨慎的表述，因为其中包含了“或许”二字，他在其他文章中采用的均是肯定语气）不仅低估了玻尔当时的知识，不同于玻尔原始论文的表述，而且从物理上讲也是不成立的。

那么究竟什么才是玻尔当时的知识，以及他本人的表述呢？我在这里引述一下，以供对历史感兴趣的读者参考，也作为对戈革先生文章的补充。玻尔的表述是这样的（该表述来自玻尔1918年初次表述对应原理的论文《论线光谱的量子论》，重点是我加的）：

没有关于定态间跃迁机制的详细理论，我们当然不能普遍地得到两个这种定态之间自发跃迁几率的严格确定法，除非各个 n 是一些大数。.....对于并不是很大的那些 n 值，在一个给定跃迁的几率和两个定态中粒子位移表示式中的傅里叶系数值之间也必定存在一种密切的联系。

在这里，玻尔明确表示，在一般情形下“当然不能普遍地得到”跃迁几率的严格确定法，在这点上连“或许”这样的含糊度都是不存在的。关于跃迁几率与相应泛频所对应的经典振幅的关系，他只表示存在“一种密切的联系”。而对于将跃迁几率视为正比于经典振幅平方的这种估算方法，他在同一篇论文中明确表示：“各个 n 值越小，这种估计当然就越不准确”，在这里，他同样没有提出或认可哪怕是“或许”意义下的普遍规律。在后来的文章中，玻尔反复强调了这种普遍联系的不存在性，比如在1921年第三届索尔维会议（Solvay conference）的发言稿中，他表

示在“跃迁的几率和运动之间得到一种简单直接的定量联系的可能性”已经被“很自然地排除了”。因此戈革对上述第四条的表述无疑是不准确的。

对应原理对今天学物理的学生已无重要性，我就不再多写了。[\(13\)](#) 有读者可能会问：既然玻尔并未提出小量子数情形下跃迁几率与经典振幅之间普遍而定量的关联，对应原理在旧量子论中为何还如此有用，以至于玻尔和玻恩对它作出了如此高的评价？这是因为，对应原理虽未能给出计算跃迁几率的普遍方法，但玻尔所说的跃迁几率与经典振幅之间的“密切的联系”包含了一些重要的定性对应，比如可以通过对经典振幅的分析确定量子跃迁为零的情形。这样他就可以导出量子跃迁的选择定则（selection rule），以及跃迁辐射的偏振性质，而这些在旧量子论时期具有极大的重要性。

五、关于互补原理

互补原理（玻尔本人通常称其为“互补性”或“互补关系”）是玻尔在1927年召开的科摩（Como）会议上提出的。它被认为是量子力学哥本哈根诠释的核心，同时也是玻尔对哲学的重大贡献。对于这一原理，很多人给予了很高的评价，也渲染了它的轰动性。比如传记作家穆尔（Ruth Moore）在其流传广泛、但本质上是外行手笔的玻尔传记中，称互补原理在科摩会议上的提出“像西北风有时搅动往常很平静的科摩湖面一样搅动了会场”。戈革先生则表示：玻尔的互补原理及由此产生的互补哲学“在学术思想界引起了轩然大波，发生了难以估计的影响”，“其影响之深远甚至远远不是相对论的影响所能比拟的”。

就连一向谦逊的玻尔本人，对互补原理也作出过相当夸张的评价。在其有生之年，玻尔曾致力于将互补性观念推广到物理学的其他分支（比如热力学），自然科学的其他领域（比如化学、生物），甚至社会科学（比如人类学、社会学、政治学）等。他曾表示，他的互补哲学能使人们的思想更加开朗，使人类的关系更加和谐，可以代替宗教，将来甚至在中小学里都会被讲授。

互补原理果真有那么重要吗？在我看来答案是否定的。起码，物理学上几乎没有任何重大进展曾经显著地得益于这一原理。这一点甚至连玻尔本人的工作也不例外。互补原理提出后，玻尔本人在物理学上最主要的工作有两项：一项是对电磁场的可测量性进行研究，它源于苏联物理学家朗道（Lev Landau）与德裔英国物理学家派尔斯（Rudolf Peierls）1931年的一些想法。另一项则是提出了原子核的液滴模型，它为早期的核裂变研究提供了帮助。这两项研究都不曾在实质意义上依赖过互补原理，虽然玻尔曾用互补原理对其中第一项研究作了一些包

装。(14)

不过，考虑到互补原理已被泛用到了非常宽广的领域，而且它作为哥本哈根诠释的核心，被很多人认为是理解量子力学所必需的，因此从正面论述互补原理的不重要性将是困难和容易引发争议的。本节不拟作这样的“强攻”，而只打算从一个非常特殊的角度出发来进行一些探讨。这个角度就是：究竟有多少人真正懂得互补原理？之所以选择这样一个角度，是因为无论我们将互补性视为哥本哈根诠释的核心，还是适用面宽广的哲学，要想确认它的重要性，确认它对后世的影响，首先应该确认它已被很多人所理解。这“很多人”虽然无需多到像玻尔所期望的那样包括未来的中小學生，起码应该包括相当比例的物理学家，以及视互补原理为研究对象的专业人士。反过来说，如果连那些人都未能理解互补原理，那么人们围绕这一原理所作的广泛探讨，就很可能是人云亦云、以讹传讹，甚至有可能是“挂着羊头卖狗肉”。一个不能被他人真正理解的观念，是很难称得上对后世有真正的影响力或重要性的。而一个物理理论的诠释，如果能理解它的人比能掌握该理论的人少得太多，其意义也是值得商榷的。

我们先来看看互补原理在科摩会议上的反响。依据当时的会议记录及物理学家们的事后回忆，我们可以确认，穆尔所述的互补原理像西北风搅动科摩湖面一样搅动会场并不是事实。玻尔的演讲结束后，玻恩、克拉默斯（Hendrik Anthony Kramers）、海森伯、费米（Enrico Fermi）及泡利等人参与了例行讨论，但那些讨论并未涉及玻尔所强调的任何重点。事实上，就连互补原理后来最热情的传播者罗森菲尔德也表示自己当时看不出、也感觉不到演讲中的任何微妙之处。与会者之一的匈牙利裔美国物理学家维格纳（Eugene Wigner）很好地总结了互补原理对会场的真正影响：“这篇演讲不会使我们中的任何人改变自己关于量子力

学的见解”。

那么，互补原理是如何从一个几乎无人在意的灰姑娘，转变成为一个让人耀眼生花的小公主呢？我个人认为，这与爱因斯坦对量子力学的诘难有很大关系。爱因斯坦的诘难在无形中给了玻尔一个像“焦点访谈”一样的重要舞台，使他能够在众人的关注之下反复重述那些此前曾被忽视的观点。由玻尔出面对爱因斯坦进行回击这一事实，以及回击过程所具有的令人回味的戏剧性，使玻尔的观点逐渐成为正统，并得到越来越多的物理学家起码是名义上的认同。[\(15\)](#)

但这种数量上的繁荣，却并不意味着有越来越多的人开始懂得了互补原理。相反，我们发现即便在互补原理的核心支持者中，对这一原理的表述也存在很大的出入。甚至连玻尔本人的论述也存在显著的易变性。这一点对于熟悉玻尔的人来说是并不陌生的。几乎所有当过玻尔助手，帮他记录过文稿的物理学家都有一个共同体会，那就是玻尔的文章大都会一改再改，而且修改的过程不一定收敛。很多文章的最后定稿并非是由于玻尔觉得满意了，而是因为交稿的时间已到，才不得不“悬崖勒马”。玻尔的科摩演讲稿就是一个典型的例子，据帮他记录文稿的克莱因回忆，文稿的准备从4月初就开始了（科摩会议是9月份召开的），每天都由玻尔口授，克莱因记录，但“到了第二天，他所口授的又全部作废”，于是一切又从头开始，“就这样度过了整个的夏天”。这样的过程有点像张三丰给张无忌演示太极剑法——两次演示的是全然不同的剑招。可惜玻尔不是张三丰，他的读者也不是张无忌。玻尔对互补原理文稿的修改甚至在科摩会议之后仍进行了相当长的时间。由这样反复修改所得的文稿尽管精细得近乎晦涩，却仍无法完整地表达玻尔的意思。爱因斯坦就曾抱怨说：“尽管花了很大的精力，我还是不能得出玻尔互补原理的确切表述。”

爱因斯坦不懂倒也罢了，他老人家本就不是量子力学阵营的。不幸的是，“革命队伍”的内部也出现了不和谐的声音。量子力学阵营少壮主力之一的狄拉克沉痛交待了自己的问题（那口气像是跟“敌人”学的）：“我没有弄懂那些思想，尽管我曾经尽了最大的努力企图弄懂它们。”爱因斯坦和玻尔的共同好友，在两人的争论中毫无保留地支持玻尔的荷兰物理学家艾伦菲斯特（Paul Ehrenfest）也在给学生的信中诉苦说：“又是玻尔那种可怕的咒语连篇，别人谁也没法总结。”爱因斯坦、狄拉克和艾伦菲斯特算是“主动自首”的。美国物理学家派斯可就没那样的觉悟了，他在自己撰写的玻尔传记中表示，对自己来说阅读玻尔的作品是并不困难的，因为自己拥有一个“不寻常的优势”，那就是曾经与玻尔本人谈论过互补性，并且“所用的小时不计其数”。派斯随后就对玻尔的互补性概念及其渊源进行了介绍和分析。限于篇幅，我就不重复他的文字了，但我要提一下戈革对派斯这番宏论的评价。简单地讲，戈革的评价中客气的部分是说派斯的分析“似是而非”，不客气的部分则斥之为“完全是胡扯”，戈革并且在“完全是胡扯”上加了着重号（相当于再踏上一只脚）。

对互补原理的另一类“理解”来自熟悉东方文化或对东方文化感兴趣的人，这方面的代表人物是日本物理学家汤川秀树及美国物理学家惠勒（John Archibald Wheeler）。这两人都表示互补原理在东方是一种很自然的思想方法，不难接受。但同样熟悉东方文化并对东方文化感兴趣的戈革并不认同这种观点，认为他们既误解了互补原理，也误解了东方文化，至多只在很有限的意义下有道理。看来戈革在这方面的眼光是相当挑剔的。那么，在被奥格·玻尔视为也许能当互补原理“亚圣”的戈革先生眼里，除玻尔本人外，还有没有真正懂得互补原理的人呢？有，但似乎只有一位，那就是“教皇的唱诗童子”罗森菲尔德。在戈革看来，罗森菲尔德才是互补原理的“亚圣”。但罗森菲尔德的“亚圣”位置却也并非众

望所归，比如奥格·玻尔就反复表示过反对，他认为罗森菲尔德太教条，而且太自以为是。

说来说去，我们还漏掉了两个重要人物，那就是海森伯和泡利。这两人都是哥本哈根学派的核心成员，他们很早就接触到玻尔的观点，并且很早就成为互补原理的支持者。⁽¹⁶⁾在这两人中，我暂未能找到泡利有关互补原理的表述（泡利是一位哲学味较淡的物理学家），因此无法评论。但海森伯曾在《物理学和哲学》（*Physics and Philosophy*）等著作中对互补原理作出过自己的表述。从他的表述来看，他对互补原理的理解与玻尔存在不小的差异。在玻尔眼里，互补原理是对量子力学作自洽理解的基础，海森伯却反其道而行之，把由互补原理所描述的局面的自洽性归因于量子力学数学体系的自洽性，从而极大地弱化了互补原理的地位。另一方面，罗森菲尔德曾表示互补原理与唯物主义的辩证法颇有互通之处（戈革对此不无微辞），但他却把海森伯归为唯心论者。因此，海森伯究竟在多大程度上皈依互补原理，是一个值得存疑的问题。说到海森伯，顺便也提一下他的学生魏茨泽克（Carl Friedrich von Weizsäcker）。这位有“量子神学家”之称的德国物理学家在玻尔70周年寿辰前仔细研读了玻尔有关互补原理的论述，并自信已理解了它的真义，但当他征求玻尔的看法时，玻尔却毫不含糊地给了他一个否定的答复。

以上我们讨论了几位主要的、与玻尔有过直接接触的物理学家对互补原理的理解（确切地说是不理解）。信奉互补原理的物理学家当然远不止我们提到的那几位，可惜多数人从未对互补原理作出过自己的表述，因此我们无法评论。不过，假如那些与玻尔有过直接接触甚至反复讨论的物理学家都不能真正理解互补原理，后世那些只能从有可能词不达意的文字叙述中学习互补原理的人无疑只会更不容易理解它。此外，

还有一点也许是很多人的共同观感，那就是绝大多数后世物理学家对量子力学的态度接近于美国物理学家费恩曼（Richard Feynman）的一个观点，那就是量子力学的核心之谜是无法通过任何“诠释”而得到真正澄清的，我们所能表述的只是量子力学是如何工作的。换句话说，经过了这么多年思考和运用，越来越多的物理学家已经意识到：这世界是一个量子世界，量子力学本身就是对这一世界最清晰的表述。今天学物理的学生如果回过头去阅读玻尔等人有关互补原理的原始论文，很可能非但不会觉得豁然开朗，反而会感到更大的困惑。而今天仍在研究量子力学诠释的人，他们走的也早已不是玻尔当年那种思辨性的老路，而是诸如退相干之类与量子力学数学体系的关系密切得多的途径。⁽¹⁷⁾

在本节的最后，我也评述一下那种认为玻尔的互补原理是对哲学的重大贡献的观点。按照与上文同样的思路，我认为互补原理对哲学的发展也无重大贡献。理由很简单：物理学家尚且很难理解玻尔互补原理的确切含义，哲学家就更没希望了。尽管哲学文献中已出现了大量讨论或引申互补原理的文章，但那些文章大都只是人云亦云或以讹传讹。玻尔曾明确表示：“我可以很合理地说，没有任何一个被称为哲学家的人真正懂得互补描述是什么意思。”玻尔被很多人视为是极富哲学气质的物理学家，但他晚年对哲学却殊少敬意。他曾在旁听完一个哲学讲座后作出了相当罕见的尖锐批评：“哲学家们所说的一切都是纯粹的胡扯。”

当然，我们无意仅凭玻尔的一两句话就否定哲学家的研究，为了看看他们到底研究了什么，我们来举一些例子。我们知道，给哲学思想贴标签是哲学家们热衷的事情之一。在有关互补原理的哲学研究中，有一大类是在讨论玻尔互补性思想的起源。较新的考察表明，几乎所有此类“研究”都是捕风捉影和牵强附会的。以最主流的研究者为例，著名的以色列科学哲学兼科学史学家雅默（Max Jammer）曾凭借丹麦神学家

克尔凯郭尔（Søren Kierkegaard）在其神学著作中谈到人生态度时用过的“飞跃”一词，就极富想象力地将之与玻尔所研究的定态间的“跃迁”联系起来，认为玻尔的思想曾受其影响。他的这种看法遭到了罗森菲尔德的批评。雅默在后来的著作中放弃了这一观点，但互补原理的“克尔凯郭尔”标签却不胫而走，被其他研究者所“吸收”，并得到了广泛的流传，这其中包括著名的美国科学哲学家霍尔顿（Gerald Holton）。同样还是雅默，又提出玻尔的互补性思想曾受到美国哲学兼心理学家詹姆士（William James）的《心理学原理》（*The Principles of Psychology*）的显著影响。可惜后来的研究给出了很强的证据，表明玻尔直到1935—1936年间才读到过詹姆士的那部著作，此时互补原理早已问世多年。类似地，玻尔在大学时的哲学教授霍夫丁（Harald Høffding）也被想当然地视为对玻尔互补性思想的提出有过实质影响。支持这一观点的研究者不仅夸大了玻尔与霍夫丁学术交往的程度及频度，将霍夫丁夸张成玻尔的“唯一哲学导师”，有人甚至还将信奉严格因果律的霍夫丁贴上反因果律的标签，以便与玻尔对因果律的放弃建立联系。至于有关互补原理的其他形形色色的哲学研究，以及将宗教与科学视为“互补”之类更等而下之的东西，我们就不再评述了。

综上所述，我认为玻尔互补原理的重要性被早期的物理学家及后来的哲学或史学家们显著夸大了——当然，上面的论述算不上是决定性的。但我想它们能说明，互补原理无论对于物理学还是哲学，其重要性起码都是有理由存疑的。[\(18\)](#)

六、关于EPR争论

史学家在阐述历史的时候，通常采用旁观者的视角，所评论或辨析的则通常只是史料或对史料的分析。但戈革先生在介绍玻尔与爱因斯坦的EPR争论时，却出现了一个引人注目的例外：他不仅对历史，而且对爱因斯坦等人的观点本身进行了评述。在1985年发表的“关于尼耳斯·玻尔思想的几点历史考察”一文中，他这样写道：

我认为，1935年的EPR论文，在出发点上存在相当大的问题。作者们的前提不但在某种程度上自相矛盾，而且大大违反了“竞赛规则”。

戈革随后举出了EPR论文中的“实在性判据”，即：

如果在不以任何方式扰动一个体系的条件下，我们能够确切地（即以等于1的几率）预言一个物理量的值，则存在物理实在的一个要素和该物理量相对应。

戈革表示：“这样一个判据，事先就把量子力学‘否定掉了一半’”，因为“按照玻尔的看法，‘不以任何方式扰动一个体系’，一般地就意味着完全不能对体系作出任何预言”，而且“‘确切地（即以等于1的几率）’作出预言则更加意味着从一开始就一般地或在原理上排除了量子力学的几率诠释”。戈革并且认为，玻尔在EPR论文问世八年之前的科摩演讲中

的一段话：

量子公设意味着，对原子现象的任何观察，都将涉及一种不可忽略的和观察仪器之间的相互作用，因此，就既不能赋予现象也不能赋予观察仪器以一种通常物理意义下的独立实在性了。

实际上已预先对EPR论文作出了回复。而EPR论文的逻辑，则等于是“一开始就假设人家是错误的”。戈革的这一观点在他的其他文章中也曾提到过。

显然，如果戈革的上述评论成立，即如果EPR论文违反了“竞赛规则”，“一开始就假设人家是错误的”，并且玻尔早在八年前就已实际上回复了EPR论文，那么不仅爱因斯坦等人的论文沦为了不值一驳的低级错误，连玻尔的眼力也成了问题。因为玻尔面对违反竞赛规则这样的巨大破绽，居然没将对方“一剑封喉”，却花了好几个月的时间来酝酿回复。据罗森菲尔德回忆：EPR论文的“突然袭击像一个晴天霹雳一样打到了我们头上，它对玻尔的影响是显而易见的”，“玻尔一听到爱因斯坦的论述，就把其他所有事情都搁下了”，然后就是“一天又一天，一周又一周”的仔细分析，直至完成回复为止。爱因斯坦等人的倾力出击，玻尔的细腻回复，果真是忽略了一个低级错误吗？我们在这里分析一下。

首先要指出的是：戈革提到的“‘确切地（即以等于1的几率）’作出预言”意味着“一般地或在原理上排除了量子力学的几率诠释”显然是不正确的。量子力学的几率诠释并不排除几率等于1的特殊情形（感兴趣的读者不妨举出一些量子力学涉及几率1的例子）。不过，这一点在戈

革的论述中并非核心，我们只是略提一笔。戈革观点的真正核心，在于他认为EPR的实在性论据等于直接否定了玻尔反复阐述过的不以任何方式扰动一个体系，就完全不能对体系作出任何预言的观点，从而等于是一开始就假定玻尔错了。在辩论中如果一方一开始就假定对方错了，后面的论述岂非变得不必要了？

为了搞清楚EPR的实在性判据是否真的等于直接否定了玻尔的观点，我们必须引述EPR论文中戈革先生没有引述的一段话，这段话其实就紧挨在实在性判据的后面（重点是我加的）：

这一判据尽管远不能穷尽判定物理实在的所有可能方式，但一旦它所要求的条件成立，它起码向我们提供了一种这样的方式。不将其视为必要，而只视为充分，这一判据是与经典及量子的实在观念相符合的。

这段话清楚地表明，EPR论文并未将他们的实在性判据视为唯一判据，而只是表示，假如该判据所要求的条件——即在不以任何方式扰动一个体系的条件下，我们能够确切地预言一个物理量的值——成立，那么存在物理实在的一个要素与该物理量相对应。EPR论文明确表示，并非只有在那样的条件下，才能定义物理实在。因此他们的实在性判据并不排斥玻尔的观点，而只是试图在那种观点之外补充一种可能性。相反，如果我们把玻尔的观点理解为EPR论文所涉及的那种补充的可能性连探讨都不能探讨，那倒反而是违背“竞赛规则”的，因为那等于一开始就假定只有玻尔才是对的。幸好，EPR论文的作者与玻尔都很懂规则，在辩论中均未做出“违规”的事情。EPR论文的重点，在于试图构造一个

能满足他们提出的实在性判据的测量过程，而玻尔的反驳，则意在说明被EPR论文以为是满足他们要求的测量过程，其实依然包含了“扰动”，从而不能被视为是满足要求的。

有关EPR争论，还有一个非常重要的地方需要提一下，那就是EPR论述作为爱因斯坦-玻尔论战的最后一轮，虽仍不曾战胜玻尔，却击中了玻尔在此前宣讲互补性时的一个盲区。那个盲区恰恰体现在戈革引述的玻尔科摩演讲的那段话里。在那段话里，玻尔强调了互补性的根源在于“对原子现象的任何观察，都将涉及一种不可忽略的和观察仪器之间的相互作用”。而在对EPR的回复中，玻尔写道（重点是原文就有的）：

当然，在刚刚考虑的这一类事例中，在测量程序的最后关键阶段是谈不上对所考虑体系的机械干扰的。但是，即使在这一阶段中，也还在本质上存在对一些条件的影响问题，那些条件确定着有关体系未来行为的预言类型。

在这里，玻尔事实上被迫作出了一次退让，引进了所谓“对一些条件的影响”这样一种纯概念性或信息性的“干扰”。这一退让虽不至于让他的观点失去自洽性，却大大削弱了它的力度。因为玻尔原本仰仗的干扰——虽不曾明说——实际上正是“机械干扰”，它与海森伯的不确定原理或作用量子的有限性一脉相承，在量子力学范围内有比较坚实的理论背景。而一旦被引进所谓“对一些条件的影响”，就部分地陷入到了哲学的甚至语义的纠缠之中。[\(19\)](#)后来研究量子力学基础的很多物理学家都注意到了玻尔的这一退让。被迫作出这一退让，有可能正是罗森菲

尔德所描述的EPR论文“像一个晴天霹雳一样打到了我们头上”的原因。EPR论证虽未能战胜玻尔，但它击碎了笼罩在玻尔互补性论述中受不确定原理或作用量子的有限性所保护的“机械干扰”这一硬壳，从而为20世纪60年代之后人们对量子力学其他诠释的研究作出了铺垫。

在结束本节前，我愿再补充几句与互补原理及EPR争论有关的评论。玻尔对EPR论文的回复无疑是借助了他的互补性观点。但有意思的是，爱因斯坦曾回忆说他们的论文发表后，很多人给他写了信，人人都说他错了，但每个人提出的理由都是不一样的。在玻尔出面回复前，海森伯也曾着手撰写了回复，他的回复并未发表，但在给泡利的信中海森伯表示他的回复与玻尔是不同的。⁽²⁰⁾我想这是又一个例子，它说明即便像海森伯这样公认的哥本哈根学派的核心成员，在实际思考问题——哪怕是思考一个互补原理可以插足的问题——时，也很少能按玻尔的互补原理出牌。至于玻尔对EPR论文的回复发表之后的情形，英国皇后大学（Queen's University）的量子力学研究者惠塔克（Andrew Whitaker）有一段评论，我在这里引述一下。惠塔克提到玻尔的回复发表之后，绝大多数物理学家都支持了玻尔，不过：

不应该假定这些物理学家中的大多数曾经深入研读过玻尔的回复，甚至不能假定他们曾经读过。对他们来说，玻尔宣布找到了爱因斯坦的错误就已经足够了。但是，等到物理学家们真的深入研读玻尔的回复时，他们的感觉常常会改变，这样的情况正在发生，而且很可能正在加速发生。因为自20世纪60年代开始，失去了干扰诠释（注：指被EPR论文击碎了的“机械干扰”）后，玻尔的观点往往显得过于抽象，且语义重于物理。

惠塔克举出的一个重要例子就是爱尔兰物理学家贝尔（John Stewart Bell），他提出的贝尔不等式（Bell's inequality）虽导致了对量子力学有利的实验结果，但他公开表示自己并不理解玻尔的观点，这些我就不进一步叙述了。

最后我想指出的是，EPR论文对物理学的实际影响远远超过了玻尔对EPR论文的回复，也远远超过了互补原理或互补哲学的影响。它不仅为后世对量子力学基础的研究作出了铺垫，而且直接间接地导致了贝尔不等式、量子纠缠（quantum entanglement）、量子信息、量子计算、量子密码等后续研究或新兴领域的出现。[\(21\)](#)从这个意义上讲，EPR论文非但不是个低级错误，而且是一项影响深远的研究，那些影响甚至不是爱因斯坦本人所能梦想到的。

七、结语

这篇纪念戈革先生的文章到这里就要结束了。本文因为有一大半内容在评述戈革先生的观点中我认为值得商榷的部分，其中包括对许多有关玻尔的过誉之词的商榷。这种商榷有可能给人一个轻视玻尔的印象。其实我对玻尔是相当欣赏的，只不过欣赏玻尔的话戈革先生基本上都说了，而本文只讨论有异议的部分，从而体现不出我对玻尔的欣赏。套用玻尔的语式来说：我欣赏玻尔的部分，其实比人们想象的还要多。[\(22\)](#)

在结束本文的时候，我愿意引用戈革先生在口述自传《正直者的困境》的结尾中所说的一段话。那是他2004年参加一套丛书的发布仪式时，想说却因时间关系没来得及说的。那段话有些地方不太通顺，大概是因为他口述时身体已很虚弱的缘故。我把它与戈革先生在《尼耳斯·玻尔集》译后记中所写的内容相似的段落归并整理了一下，作为本文的结尾：

人们都认为这些书不会有读者，而且和现在的生活距离很远，但我想举一个例子。20世纪物理学家中有一个脾气最坏的苏联科学家，叫做朗道，他是个天才，但他总是故意惹人生气，对世界十分挑剔，甚至讨厌别人刮胡子。他一生周游世界，旅行的范围远远超过了一个普通苏联人的行动边界。朗道年轻时思想很“左倾”，看不起“资产阶级的知识分子”。据说有一次，他在某学术中心的图书室中闲逛时感到无聊，便随手从书架上抽下一本书来说：“让我看看资产阶级的知识分子在讲些什么。”他开始看那本书，据说那是拉普拉斯

论文集中的一卷。他看着看着，脸上那种轻蔑的微笑消失了。又过了一会，他找了个位子坐下来，埋头读起那本书来，狂躁的朗道完全投入到了书中的世界里……这个故事给我留下了深刻的印象。我常常幻想，不知在我死后的若干年中，我的译本能否遇到朗道这样的知音人。这真是一种十分可笑而又可怜的想法。但我想，即使最终也没有人去读它，它存在那里，就已经改变了中国。

最后，我愿向戈革先生表达一个后辈最由衷的敬意及谢意。在具体的史学问题上人们或许各有见解，但史学文献却是共同并且永久的财富。感谢戈革先生用一生的辛勤笔耕，为中国科学史学界提供了如此丰富的文献。

愿戈革先生安息！

参考文献

[1] 戈革. 史情室文帚（上、下）[M]. 北京：中国工人出版社，1999.

[2] 戈革. 红苕论猫 [M]. 2001.

[3] 戈革. 关于书房 [M]. 见董宁文. 我的书房.长沙：岳麓书院，2005.

[4] 戈革. 《尼耳斯·玻尔集》译后记 [J]. 科学文化评论，2007，5.

[5] 戈革口述. 邹波，李翔整理. 正直者的困境，2006.

[6] 玻尔. 尼耳斯·玻尔哲学文选 [M]. 戈革，译. 北京：商务印书馆，1999.

[7] 玻尔. 尼耳斯·玻尔集（第三卷）[M]. 戈革，译. 北京：科学出版社，1990.

[8] 玻尔. 尼耳斯·玻尔集（第六卷）[M]. 戈革，译. 北京：科学出版社，1991.

[9] 玻尔. 尼耳斯·玻尔集（第七卷）[M]. 戈革，译. 北京：科学出版社，1998.

[10] 派斯. 尼耳斯·玻尔传 [M]. 戈革，译. 北京：商务印书馆，2000.

[11] 否尔霍耳特. 尼耳斯·玻尔的哲学背景 [M] . 戈革, 译. 北京: 科学出版社, 1993.

[12] 玻姆. 量子理论 [M] . 侯德彭, 译. 北京: 商务印书馆, 1982.

[13] 雅默. 量子力学的哲学 [M] . 秦克诚, 译. 北京: 商务印书馆, 1989.

[14] 卡西第. 海森伯传 [M] . 戈革, 译. 北京: 商务印书馆, 2002.

[15] A hero of history of science: Niels Bohr's Chinese translator [J] . AIP History Newsletter, 2001, XXXIII (2) .

[16] Bohr N. Phys. Rev., 1935, 48: 696.

[17] d'Espagnat B. Conceptual Foundations of Quantum Mechanics [M] . New York: Perseus Books Publishing, LLC., 1999.

[18] Einstein A, Podolsky B, and Rosen N. Phys. Rev, 1935, 47: 777.

[19] Feynman R. The Feynman Lectures on Physics (Vol III) [M] . Boston: Addison-Wesley Publishing Company, 1965.

[20] Heisenberg W. Physics and Philosophy [M] . New York: Harper Perennial Modern Classics, 2007.

[21] Waerden B L. Sources of Quantum Mechanics [M] . New

York: Dover Publications, Inc. 1968.

[22] Whitaker A. Einstein, Bohr and the Quantum Dilemma [M]. Cambridge: Cambridge University Press, 2006.

二零零八年五月四日写于纽约

二零一三年十月二十六日最新修订

(1) 2012年,《尼耳斯·玻尔集》中译本全部12卷终于被华东师范大学出版社出版。

(2) 这段话写于1990年,到他去世时,他研究玻尔已近半个世纪。

(3) 这一点是从他发表于2001年的文章《红苕论猫》中的文字“吾母逝时,吾方身在难中”加以推测的,倘若有误,欢迎知情的读者不吝指正。

(4) 不过在这一点上,戈革先生的叙述在我看来也有所欠缺,我将在本文的第四节中加以评述。

(5) 戈革也曾表述过爱因斯坦比玻尔更伟大的观点。但他表述那种观点时大都带有一个起中和作用的补充说明。比如他曾表示“爱因斯坦应该是20世纪最伟大的物理学家”,但随即补充说“因为他的名气最大”。我们都知道,名气这东西来源是很复杂的,如果伟大仅仅是因为名气大,那这种伟大完全可能是带水分的。又比如他曾表示“爱因斯坦只有牛顿配与他相提并论”,“从高度上而言,玻尔也不及他”,但随即又表示“爱因斯坦是高山,玻尔是很大的山,他们俩谁更伟大很难讲”。另一方面,当他表示玻尔比爱因斯坦更伟大时,通常讲得很干脆,不拖泥带水。

(6) 上面的叙述比较简化,这里补充几点:(1)“当电子从一个定态过渡到另一个定态时,它怎么决定将以什么频率来振动”是卢瑟福(Ernest Rutherford)在玻尔的论文发表之前写信提出的问题,玻尔在回信中未予回答。(2)玻尔在建立原子模型时,也引进过重要假设,比如认为处于基态的电子不辐射电磁波,但那些并不是人们通常所说的比相对论更深刻的观念性突破,即对经典决定论与经典实在观的突破。(3)波粒二象性的确立也有来自理论——比如爱因斯坦光子假说——的影响。(4)除散射实验外,波函数几率诠释的提出也曾受到爱因斯坦提出的电磁波与光子关系的影响。

(7) 这里也补充几点:(1)在狭义相对论创立之前已有很多实验积累,但它们并未对爱因斯坦的研究产生重大影响。倒是洛伦兹(Hendrik Lorentz)等人在那些实验引导下所做的努力,可以归类为实验结果与理论框架先行,从而更接近于量子理论的发展模式,可惜并未成功。(2)广义相对论三大经典验证之一的水星近日点反常进动作为一个观测结果,早在广义相对论创立之前就被发现,但并未对爱因斯坦的研究产生任何影响(它只是在广义相对论研究接近尾声时起到了检验作用)。

(8) 而且爱因斯坦一举奠定了直至今今天依然适用的相对论基本框架，其工作结果之完善，战场打扫之干净，使得他有生之年，基本上无人能在同一领域中作出重大贡献，其影响力无疑进一步打了折扣。

(9) 关于玻尔与年轻人的合作，奥地利物理学家韦斯科夫（Victor Weisskopf）在其自传《洞察的乐趣》（The Joy of Insight）中有几段有趣的回忆。他说跟玻尔合作的人常被戏称为“受害人”（victim）。合作的过程往往是这样的：一旦玻尔心血来潮要讨论时，“受害人”就被叫进玻尔的办公室，在一张桌子旁坐好。玻尔自己则在办公室里走动，每隔几分钟绕“受害人”走一圈。在走动的过程中，玻尔一边思考一边表述自己的想法。“受害人”如果听不懂或觉得不够清楚就需要大声说出来，以便玻尔改进表述。口述论文时玻尔通常会使用极长的德文句子，要求“受害人”记录下来，玻尔还常常修正自己的长句（“受害人”当然也就得重新记录）。韦斯科夫当“受害人”时曾希望玻尔将长句拆成短句，未获同意；韦斯科夫还曾希望自己能站起来走走，也未获同意。玻尔的规则是“只能有一个人走动”（这个人是谁当然不言而喻），因此韦斯科夫只能连续几天“罚坐”。当然，韦斯科夫表示当“受害人”其实是难得而荣幸的机会，能亲眼目睹玻尔如何做研究。不过，更“厉害”的年轻人有时不那么看。比如狄拉克只当了半小时的“受害人”就“拍案而起”，对玻尔说：“我在想，学语文时是否有人告诫过您不要在完全想好之前就把一句话写下来？”（这句话太长，不像传说中的狄拉克风格，不过既然是韦斯科夫回忆录中提到的，就姑妄听之吧。）

(10) 零星理由的另一个例子，是引述若干物理学家对玻尔的推崇。这种选择性的引述只能为文字增色，却不足以作为论述依据。因为人们可以找到更多推崇爱因斯坦的物理学家，其中甚至有可能包括推崇过玻尔的那几位。

(11) 当然，玻尔在酝酿论文的过程中还是会进行必要的计算或估算的，但其复杂度显然无法与创立量子力学所要求的相比。

(12) 那时玻尔还不曾采用“对应原理”这一名称，后者是从1920年起才使用的。

(13) 考虑到对应原理原始定义之过时，现代误解之广泛，不排除人们将错就错，把误解吸纳为约定成俗的新定义的可能性。

(14) 玻尔有关电磁场量可测量性的研究发表于1933年，是与罗森菲尔德合作的。玻尔的这项研究是他中后期罕见的一项涉及大量数学计算的研究，但其中的数学计算大都是由罗森菲尔德完成的。

(15) 需要指出的是，除了后文会提到的EPR争论外，玻尔对爱因斯坦其他几次诘难的回复对互补原理并无实质依赖。另外值得一提的是，爱因斯坦对量子力学的诘难，以及玻尔挺身而出所作的回复，不仅有助于确立玻尔观点的正统性，而且在很多人心中树立了他与爱因斯坦双雄并立的印象，对于奠定玻尔在量子力学史上的领袖地位也有一定助益。

(16) 海森伯与玻尔曾存在过重大的观点分歧，经过长时间筋疲力尽的讨论后，海森伯终于表示接受玻尔的观点。泡利对互补原理的接受在总体上从未反复，只在细节上有过轻微的波折。1928年1月，当泡利得知玻尔在互补原理的表述上换了一份文稿后，曾表示“您换了一份底稿也使我实在高兴。事实上，在过了一段时间以后，我对旧稿并不特别喜欢了”。

(17) 值得一提的是，罗森菲尔德对互补原理的支持似乎是比较极端的，甚至曾反对将互补原理仅仅视为量子力学的诠释，而认为互补原理是量子力学不可分割的一部分。按照他的这个看法，互补原理提出之后量子力学的诠释问题已不复存在，后世在这方面的研究全都是在研究伪问题。看来奥格·玻尔批评罗森菲尔德太教条、太自以为是是很有道理的。

(18) 最后补充一点：有关互补原理的诸多误解之一，就是把量子力学对传统观念的几乎所有冲击都归结于这一原理。其实，如果玻尔没有提出互补原理，量子力学对传统观念的冲击并不会减弱。一个很好的例子就是爱因斯坦，他从未搞懂过互补原理，但却依然感到量子力学深

深冲击了他的观念（他的这种感觉早在互补原理提出之前就有了），从而一再提出诘难（他的诘难没有一个是直接针对互补原理的）。

(19) 在玻尔引进这种“对一些条件的影响”之后，EPR的实在性判据倒的确可以被认为——从玻尔的新立场看——是“直接否定了玻尔的观点”，只不过这时的冲突乃是玻尔修改立场所导致的，从而谈不上是违反“竞赛规则”。

(20) 海森保的文稿后来被收录于他与泡利的通信集中，可惜他们的通信集似乎没有英译本。

(21) 需要指出的是，美国物理学家玻姆对EPR论述的简化对这些后续发展也有非常重要的贡献。

(22) 这是套用玻尔的一个有名的句式：“我们赞同的比你想象的还要多得多（We agree much more than you think）。 ”

玻尔的错误⁽¹⁾

一、引言

我曾经翻译过美国物理学家温伯格（Steven Weinberg, 1933— ）的一篇文章，标题是《爱因斯坦的错误》（*Einstein's Mistake*）。对于喜爱物理学史的读者来说，那篇文章列举的错误也许都是“熟面孔”，因为“爱因斯坦的错误”是一个很吸引人的话题，很多人都谈论过。从某种意义上讲，判断一位科学家是否伟大的一个另类但很管用的指标，就是看他（她）是否连所犯的错误都能吸引人们持久而广泛的兴趣。如果是，那就几乎可以断定为是伟大的科学家。在20世纪的物理学家中，爱因斯坦（Albert Einstein, 1879—1955年）无疑是那样的人物，这是对他“首席物理学家”地位的很好的佐证。

读者也许会问，温伯格列举的爱因斯坦的错误既然都是“熟面孔”，我为什么还要翻译呢？答案之一是温伯格在介绍那些错误之余还阐述了一些值得回味的观点。这种观点的一个例子，是他所说的“领袖科学家所犯的错误往往比他们的成功更能让人洞察他们那个时代的精神与背景”。不过，假如我们把目光从“爱因斯坦的错误”这一热门话题上挪开，投向一个更开阔的视野，那么在我看来最能印证这句话的与其说是爱因斯坦的错



误，不如说是另一位领袖科学家玻尔 丹麦物理学家玻尔

（Niels Bohr, 1885—1962年）的错

误。因为爱因斯坦的错误大都具有个人色彩，而且当他犯下某些错误时，他往往已处在“孤家寡人”的位置上，从而已不再是能让人洞察“那个时代的精神与背景”的最好例子。而玻尔的错误虽然远不如爱因斯坦的错误那样出名，甚至可以说是冷僻话题，但他在犯错时却是比爱因斯坦更具“那个时代的精神与背景”的领袖科学家，他的错误也因此要比爱因斯坦的错误更能让人洞察“那个时代的精神与背景”。

本文就来谈谈“玻尔的错误”这一冷僻话题。对于玻尔的粉丝来说，最好能从上文所述的另类指标的角度来解读本文，即本文与其说是在揭玻尔之短，不如说是在往玻尔作为伟大科学家的那个另类指标上添加砝码。

玻尔究竟犯过多少错误？似乎没有人罗列过，不过可以肯定的是，他犯错的数量与类型都远不如爱因斯坦那样“丰富多彩”。当然，这与其说是他在避免犯错方面比爱因斯坦更高明，不如说是因为他的研究领域远不如爱因斯坦的宽广，从而犯错的土壤远不如爱因斯坦的肥沃。在玻尔所犯的误差中，值得介绍的在我看来只有一个，那就是他对能量动量守恒的放弃。这个错误他几乎坚持了十年，且在两个独立情形下犯过，是他所犯误差中最显著，并且也最能让人洞察“那个时代的精神与背景”的。

二、玻尔的第一次错误：**BKS**理论

说起来有些出人意料，玻尔作为量子论的著名先驱，对开启了量子时代的“光量子”概念却长期抱有比较暧昧的态度，那态度说成是“拒绝”可能有些夸张，说成是“不热衷”则可能还不够分量，恰当的说法也许是“消极”吧，即“能不用就不用”，甚至不惜为不用而付出一定的代价。⁽²⁾他的这种态度几乎一直维持到了旧量子论时期的终结。有人也许会把这种态度本身也视为错误，但平心而论，这种态度对于一位量子论先驱来说虽有些出人意料，在当时的情况下却算不上错误，因为当时的实验对光量子概念的支持尚未达到判决性的程度。不过玻尔的消极态度虽不是错误，他为这种态度所付出的代价却比绝大多数同时代人大得多，大到了变成错误的程度，因为他放弃了能量动量的守恒。

玻尔这一错误的出现时机很有戏剧性。喜欢物理学史的读者大都知道，对光量子概念的早期支持主要来自两组实验现象：较早的一组是光电效应（photoelectric effect），是它促使爱因斯坦提出了光量子概念；较晚的一组则是康普顿效应（Compton effect）。玻尔因出于对光量子概念的消极态度而放弃能量动量的守恒，恰恰发生在美国物理学家康普顿（Arthur Compton, 1892—1962年）发现了康普顿效应之后，堪称是“顶风作案”。

当然，这“顶风作案”绝不是因为玻尔藐视实验，而是有具体的理由及机缘的。

那理由是：当康普顿发布他的实验结果时（1923年），他所观测到的现象仅仅是X射线在被物质散射后，其波长（从而频率）发生了变化。这一结果虽然与经典电磁理论相矛盾，⁽³⁾而与光量子概念相一致

——荷兰物理学家德拜（Peter Debye, 1884—1966年）与康普顿本人都用光量子概念诠释了实验结果，却不足以对后者构成判决性的支持，即证实光量子本身的实在性。因为康普顿当时的实验还比较粗糙，无法对基元过程（即单个光量子与电子的相互作用）进行观测，从而只能在平均意义上证实光量子概念的效力。事实上，不仅康普顿效应如此，早期的光电效应也是如此，只能在平均意义上证实光量子概念的效力。正是这一美中不足给玻尔提供了“作案机会”。

当康普顿在美国取得那些进展时，身在欧洲但消息灵通的玻尔几乎第一时间就获悉了消息。1923年底，他亲自访问了美国，与康普顿等人进行了直接交流。1924年初，在给英国物理学家卢瑟福（Ernest Rutherford, 1871—1937年）的信中，玻尔叙述了自己对美国之行的观感。那封信很清楚地显示出他对波动理论的青睐，以及对光量子观念的消极态度。他表示，康普顿的光量子诠释对于像他这样“视波动理论为信条的人”来说“简直是可怕的”。⁽⁴⁾

以上是理由，下面说说机缘。那机缘是：一位“带艺投师”的美国小伙子斯莱特（John C. Slater, 1900—1976年）恰好于1923年底造访了玻尔的大本营哥本哈根。在那次访问中，他带来的一个有关辐射与物质相互作用的有趣想法引起了玻尔及其合作者荷兰物理学家克拉默斯（Hendrik Anthony Kramers, 1894—1952年）的兴趣。三人很快就合写了一篇题为《辐射的量子理论》的论文。那篇论文所提出的理论后来被冠以三人的姓氏首字母而称为了BKS理论（那论文本身则被称为了BKS论文）。正是那篇论文，记录了玻尔的错误。

BKS理论主要包含三个核心想法，分别来自三位作者。第一个想法是所谓的“虚辐射场”（virtual field of radiation），它被认为是不同原子间的一种联系，并具有诱发量子跃迁的功能，这是斯莱特的贡献；第二

个想法是放弃建立在光量子概念之上的不同原子对辐射的吸收与发射间的因果联系，这个表述得有些含糊的想法是克拉默斯的点子；第三个想法则是放弃基元过程中的能量动量守恒，而将之弱化为一个统计性的定律，这是玻尔的馊主意。玻尔提出这种馊主意的“作案动机”是什么呢？是想调和原子能级变化的不连续性与波动理论所要求的辐射能量变化的连续性之间的矛盾，归根到底，还是对光量子观念的消极态度在作祟。

在BKS理论的三大核心想法中，斯莱特的想法是核心中的核心，可以说，没有斯莱特对哥本哈根的造访，就不会有玻尔对这一理论的掺和，这是我们将之称为“机缘”的原因。但如果因此就把玻尔的错误归咎于斯莱特，则是对后者的明显不公。因为早在与斯莱特建立任何联系之前的1919年，玻尔就在与同事的通信中多次提到了放弃能量动量守恒的想法。事实上，玻尔将对能量动量守恒的放弃塞进BKS理论是斯莱特所反对的，在后者的原始想法中不仅没有放弃能量动量守恒，甚至还为光量子概念留出了位置，只是由于敌不过玻尔的滔滔雄辩才同意了玻尔的观点。后来当BKS理论被实验证伪后，玻尔向斯莱特表示了歉意，而斯莱特当时虽然客气地表示了不介意，时隔近四十年后（那时玻尔已经去世）却在接受访谈时表达了对玻尔的强烈不满，甚至用上了“我对玻尔先生不曾有过任何敬意，因为我在哥本哈根度过了一段可怕的日子”那样罕见的语气。不仅如此，BKS论文的另一位作者克拉默斯一度也是反对玻尔对光量子的看法及对能量动量守恒的放弃的。可惜他也不是玻尔的对手，被后者日夜不停的辩论累垮，送进了病房，最终也同意了玻尔的观点。

三、BKS理论的放弃

BKS理论的出炉引起了很多物理学家的关注。在玻尔的声望及雄辩能力影响下，很多人投了诚。比如不久之后将创立矩阵力学的海森伯（Werner Heisenberg, 1901—1976年）一开始虽持有怀疑立场，表示从这一理论中“看不到实质的进展”，但在访问了哥本哈根之后，却被玻尔“洗了脑”，自己投诚不算，还将玻尔的观点“布道”给了玻恩（Max Born, 1882—1970年）。玻恩在稍后给玻尔的信中介绍了“布道”的结果：“我愿意告诉您，我对您在辐射理论问题上的新做法感到多么高兴”，“虽然我只听了海森伯的口头简报，但我很相信您的新理论是正确的，并且从某种意义上讲是对这些问题的最终答案”。

不久之后将创立波动力学的薛定谔（Erwin Schrödinger, 1887—1961年）也表示了大体上的认同。他在给玻尔的信中不仅对其新近的观点“极其同情”，还表示自己长期以来对此类想法也一直很有兴趣。不过有意思的是，他随后顺着玻尔观点所作的分析却得到了诸如能量的不确定性，单一体系的不稳定性等推论，很像是在用归谬法反驳玻尔，以至于《尼耳斯·玻尔集》（*Niels Bohr Collected Works*）第五卷的主编斯陶耳岑堡（Klaus Stolzenburg）在介绍这段历史时干脆将薛定谔的观点称为是“热力学上的反驳”。考虑到薛定谔后来几乎是仅次于爱因斯坦的反哥本哈根的人物，他对玻尔观点的“极其同情”是真正的“爱心”，还是圆滑的“外交”，倒是有点难说了。

就连泡利（Wolfgang Pauli, 1900—1958年）这位素以思维犀利、批评尖刻著称，并享有“物理学的良心”及“上帝的鞭子”美誉的年轻高手也没能抵挡住玻尔的雄辩，在访问哥本哈根时步海森伯的后尘向玻尔投了诚。但泡利毕竟是泡利，不像海森伯和玻恩那么好糊弄，在离开哥本

哈根后不久就幡然醒悟，在给玻尔的信中宣布：“您当时成功地堵住了我那强烈反对这一诠释的科学良心。但这只能持续一小段时间，……我今天又成为完全反对这一诠释的物理学家了……。”泡利的这一反对立场此后再无丝毫动摇，一直持续到BKS理论破产为止，以至于玻尔后来表示泡利“长期以来就是对我们的‘哥本哈根叛乱’不表同情的”。

连泡利都没能挡住玻尔的影响（虽然只是暂时的），可见玻尔的领袖科学家地位不是吹的。但玻尔的影响虽大，有一个人却自始至终都旗帜鲜明地反对BKS理论，这个人是谁呢？大家应该能猜到，是的，他就是玻尔的老对手、超级大腕爱因斯坦。他在获悉了BKS理论之后，在给玻恩的信中毫不含糊地表示了反对，并写下了一段后来很出名的话，那就是假如BKS那样的理论是正确的话，“我宁愿去当一个修鞋匠，甚至赌场的雇员，也不愿做物理学家”。爱因斯坦并且提出了很多具体的反对意见，泡利在离开了哥本哈根后之所以这么快就“反水”，除了他自己的思维犀利外，爱因斯坦的观点也起了一定的鼓舞作用（泡利在“反水”前曾与爱因斯坦讨论过BKS理论）。

最终对BKS理论构成重击的则是实验判决。这一判决来得很快，距离BKS论文的发表仅仅过了一个月左右，德国物理学家玻特（Walther Bothe, 1891—1957年）和盖革（Hans Geiger, 1882—1945年）就完成了一篇重要论文，对康普顿效应进行了细致研究，其初步结果对BKS理论很不利。次年（1925年）4月，他们又发布了改进的结果。与此同时，康普顿本人及其合作者也发布了更精密的研究结果。这些结果表明康普顿散射中反冲电子与散射光的出现存在明显的同时性及角度相关性，这是最初的康普顿实验因未能揭示而给玻尔等人以可乘之机的细节。这一细节与BKS理论是完全矛盾的，因为后者所预言的散射光的发射在时间及方向上都具有随机性，与反冲电子之间不存在显著的同时性

及角度相关性。那些更精密的研究还直接证实了基元过程中的能量动量守恒，从而给玻尔的观点判下了死刑。

在无可辩驳的观测事实面前，玻尔终于投降了。1925年4月21日，他在给英国物理学家福勒（**Ralph H. Fowler**, 1889—1944年）的信中承认“除了为我们的革命性努力举行一个尽可能光荣的葬礼外，已经没别的事情可做了”。在给德国物理学家弗兰克（**James Franck**, 1882—1964年）的信中，他则不无伤感地表示“此刻我觉得很不幸，而且不知如何是好。我只想向你引述瑞利勋爵的话‘某些最了解我的人认为我应该比现在更自信，也许他们是对的’”。该信的落款是“您的不幸的玻尔”。

四、玻尔的第二次错误：科学革命综合症

玻尔的第一次错误就这样“游戏结束”了。对于这次错误，美国物理学家派斯（Abraham Pais，1918—2000年）有一个评价，那就是“它比任何别的贡献更好地显示了最好的物理学家所体验到的那种紧张和混乱”。从这个评价上看，它确实很能让人洞察“那个时代的精神与背景”。不过它的这一功能与玻尔的第二次错误相比还是要逊色一筹。

玻尔第二次错误的类型很多人也许都不会想到，因为那居然是重犯——第一次错误的重犯。这种重犯的情形发生在像玻尔这样的大物理学家身上是有值得探究的原因的。

20世纪前半叶物理学史的最突出特点无疑是发生了科学革命——相对论革命及量子力学革命。如果除此之外还要归纳什么东西的话，我觉得很值得关注的是这样一种现象，那就是在经历了像量子力学革命那样激动人心的科学革命后，那一代的某些物理学家似乎产生了强烈的“科学革命情结”。一般认为，科学革命是年轻人的专长，科学史的发展也基本佐证了这一点，但经历过量子力学革命的某些物理学家却似乎是例外，在面对新挑战时，已不再年轻的他们往往比年轻人更青睐于用科学革命的思路去解决问题，以至于一有风吹草动就怀疑理论基础需做重大变革。若要给这种现象取个名称的话，我想不妨称为“科学革命综合症”。

在“科学革命综合症”的“病人”中，除玻尔外，狄拉克（Paul Dirac，1902—1984年）是很典型的一位，当量子场论被发散问题所困扰时，他一再表示应该对理论基础进行变革，而且那变革将会像从玻尔的旧量子论过渡到量子力学那样剧烈。与他那激情燃烧的“革命情怀”完全相反，

像费恩曼（Richard Feynman，1918—1988年）那样的年轻小将们却反而很实用主义地采取了非革命性的手段（重整化）来解决问题。海森伯是另一个例子，当量子场论遇到困难时，他也认为局势类似于当年的旧量子论时期，主张建立“新物理”，沿着那样的思路，他陷入了自己的死胡同——非线性旋量理论。不止一位那一代物理学家共同患有的这种“科学革命综合症”，是我认为玻尔的错误比爱因斯坦的错误更能让人洞察“那个时代的精神与背景”的主要原因。⁽⁵⁾

回到玻尔的第二次错误上来。BKS理论虽然寿终正寝了，但玻尔的“科学革命综合症”促使他在1929年左右又重新向能量动量的守恒发起了冲击，试图解决一个当时尚无答案的问题： β 衰变中的能量问题。

当时所谓的 β 衰变，是指核子通过发射电子而进行的衰变。而所谓 β 衰变中的能量问题，是指原本被认为是由衰变前后的核子状态所确定的电子能量，实际上却被发现是连续分布的，从而与能量守恒定律相矛盾。这一问题使玻尔重新提出了能量动量在基元过程中有可能不守恒的观点。不过，他再次提出这一主张，实验虽是直接契机，“科学革命综合症”的作用却也不容小觑。在经历了量子力学革命后，一个很容易被提出的问题就是：从宏观尺度进入原子尺度时我们经历了量子力学革命，从原子尺度进入到更细微的原子核尺度时，是否要经历另一次科学革命？在玻尔看来这一问题的答案是肯定的。⁽⁶⁾他并且将之与自己不久前提出的互补原理联系在了一起，认为“我们在这里遇到了对描述自然现象的习惯模式进行更激烈变更的必要性，这种变更意味着对互补性观点的进一步推广”。

1929年，玻尔将自己的观点写成一篇题为《 β 射线谱和能量守恒》的短文寄给了泡利。在那篇短文中，他不仅提出了 β 衰变中能量动量不守恒的可能性，而且还设想这种不守恒性或许有助于解释当时尚未盖棺

论定的太阳的发光之谜。不过有了上一次的前车之鉴，在给泡利的信中，玻尔谦虚地表示“我将很乐意听取您有关所有这些的看法，无论您觉得适宜用多么温和或多么严厉的语气来表达”。泡利没有辜负玻尔的信任，看完之后给出了很“温和”的评价：“我必须说它几乎没给我带来任何满足。”在作了若干技术性批评后，他的最终建议是：“让这篇短文先休息一长段时间，并让星星安静地照耀它吧。”

也许是泡利反对的缘故，玻尔最终没有发表那篇短文。但他并未就此死心，在接下来的几年间，他在信件、会议讨论及公开演讲中不止一次地提到核物理中能量动量不守恒的可能性。而泡利本人则于1930年提出了能量问题的正解，那就是 β 衰变在发射电子的同时还发射了一种看不见的中性粒子，是它带走了一部分能量，使其余部分看起来不守恒了。泡利提议的中性粒子就是我们如今所说的中微子（确切地说是反电子中微子），可惜对它的实验证实是二十多年后的事，因而未能及时终止玻尔的第二次错误。

玻尔直到1936年才放弃能量动量不守恒的提议。那时虽然中微子仍未被观测到，但意大利物理学家费米（Enrico Fermi, 1901—1954年）在中微子假设基础上建立起来的四费米子相互作用（four-fermion interaction）理论得到了很好的实验支持，使玻尔觉得问题已基本得到了解决。当玻尔最终从错误中走出来时，他不仅自己走出来了，而且对来自其他人的类似想法产生了抵御能力。如前所述，当量子场论被发散问题所困扰时，狄拉克显示出了“科学革命综合症”的“症状”，他的一个早期提议很接近玻尔曾经犯过的错误，即认为能量动量有可能不守恒，他表示“物理学目前已面临了不得不对基础进行激烈变更的前景，这种变更包括了放弃我们依赖最深的某些原理（比如能量动量的守恒），而代之以BKS或与之类似的理论”。

但此时的玻尔已不再为这类提议所动了，他对狄拉克认真看待BKS论文的做法表示了“深受感动”，但对狄拉克的观点本身却表示“一点都不满意”，他并且明确宣布那篇旧作（即BKS论文）“已经完成了自己的使命”。

五、结语

玻尔的错误就聊到这里了，毋庸置疑的是，我们对所谓“错误”的判断是建立在对物理原理的当前理解之上的。我们将玻尔对能量动量守恒的放弃归为错误，并不意味着认定能量动量守恒绝不可能在未来某一天被发现遭到破坏。不过即便有一天它被发现遭到破坏，那破坏也绝不可能如玻尔设想的那样显著，更不可能如玻尔设想的那样去解决昔日那些问题。从这个意义上讲，即便有那样一天，玻尔的错误也依然是错误。

同样毋庸置疑的是，无论玻尔的错误还是爱因斯坦的错误，都无损于他们作为伟大科学家的地位，也无损于我们对他们的敬意。科学史上几乎没有哪位伟大的科学家是从不犯错的，真正不犯错的往往反而是小角色。比如本文作者也发表过论文，那些论文也许不存在技术性错误，但与玻尔或爱因斯坦的论文相比，它们充其量只是小习题，出错的可能性虽小，有价值的可能性却更小。用一个也许不太恰当的比喻来说：在宁静小湖畔行走的人或许能不湿脚，在汹涌海浪前搏击的人却必然会沾水，伟大的科学家是后者而不是前者。

参考文献

[1] Bohr N. Collected Works (vol 2) [M] . Amsterdam: North-Holland Physics Publishing, 1981.

[2] Bohr N. Collected Works (vol 5) [M] . Amsterdam: North-Holland Physics Publishing, 1984.

[3] Bohr N. Collected Works (vol 9) [M] . Amsterdam: North-Holland Physics Publishing, 1986.

[4] Pais A. Inward Bound: Of Matter and Forces in the Physical World [M] . Oxford: Oxford University Press, 1988.

[5] Schweber S. QED and the Men Who Made It [M] . Princeton: Princeton University Press, 1994.

[6] 爱因斯坦. 爱因斯坦文集（第二卷）[M] . 范岱年，等，译. 北京：商务印书馆，1979.

[7] 派斯. 尼耳斯·玻尔传 [M] . 戈革，译. 北京：商务印书馆，2001.

二零一一年六月十八日写于纽约

(1) 本文章发表于《现代物理知识》2013年第5期（中国科学院高能物理研究所）。

(2) 有读者可能会问：玻尔早在著名的“三部曲”（玻尔原子模型）中就引用了光量子的能量

公式 $E=h\nu$ ，怎么能说是对光量子概念持消极态度呢？这是因为，玻尔虽然引用了 $E=h\nu$ ，却基本上是作为单纯的能量公式来运用的。他也用“量子”这一术语，却也基本上只是在“能量量子”的意义上使用的，而非像爱因斯坦那样引入具有粒子性的“光量子”概念。在他的“三部曲”中，电子跃迁所发射的是电磁波意义上的单频辐射，只是其能量由 $E=h\nu$ 给出而已。他在其他场合的一些论述及通信（其中有些在本文中将被引述）也佐证了这一点。

(3) 经典电磁理论对电磁波散射的描述是：带电粒子（通常是电子）在入射电磁波的作用下发生受迫振荡，并因此而发射电磁波（即散射波）。按照这种描述，电磁波在被物质散射后，其频率（从而波长）是不会改变的。

(4) 有趣的是，同为光量子诠释，玻尔对光电效应的光量子诠释似乎从未作出过那样的评价，其原因在我看来是因为康普顿的光量子诠释更强调了基元过程，从而赋予了光量子概念更明确的实在性。当然，也不排除是因为康普顿当时对玻尔来说还只是Mr. Nobody。

(5) 有读者可能会问：爱因斯坦是否也患有“科学革命综合症”？从某种意义上讲，答案是肯定的，温伯格所说的爱因斯坦晚年“变成他自己成就的囚徒”就可以被视为是某种程度的“症状”。但是，那些使爱因斯坦成为“囚徒”的成就要么是他的“独角戏”（比如相对论），要么是他所继承的经典物理观念（比如决定论），以“那个时代的精神与背景”而论显然都不如玻尔的“症状”有代表性。

(6) 当时由于中子尚未被发现，原子核理论除了面临 β 衰变中的能量问题外，还被诸如自旋、统计、磁矩、电子如何存在于核内等一系列其他问题所困扰，那些问题都加重了事态的严重性，也加重了想象中那“科学革命”的紧迫性。

希尔伯特与广义相对论场方程⁽¹⁾

一、引言

众所周知，20世纪最著名的物理学家是爱因斯坦（Albert Einstein，1879—1955年），爱因斯坦最著名的成就是广义相对论。关于广义相对论的提出，爱因斯坦的晚年合作者、波兰物理学家英菲尔德（Leopold Infeld，1898—1968年）在《爱因斯坦：他的工作及对我们世界的影响》（*Albert Einstein: His Work and Its Influence on Our World*）一书中曾经记述过一段有趣的对话：

.....我曾对爱因斯坦说：“无论您是否提出，我相信狭义相对论的问世都不会有什么延误，因为时机已经成熟了。”爱因斯坦回答说：“是的，这没错，但广义相对论的情形不是这样，我怀疑直到现在也未必会有人提出。”

英菲尔德并且评论说，这一回答很好地表述了爱因斯坦在广义相对论发展史上所扮演的角色。

确实，在20世纪前25年所发生的那场影响深远的物理学革命中，狭义相对论、广义相对论和量子力学这三大理论的提出可谓是各有特色：狭义相对论是水到渠成、瓜熟蒂落，量子力学则是群雄并起、共襄盛举——这两者虽很不相同，但似乎都是“离了谁地球照样转”。唯有广义相对论，几乎是爱因斯坦“一个人的战斗”，是没有爱因斯坦就没有广义相对论。这一点不仅有爱因斯坦本人与英菲尔德的上述对话作脚注，也是

很多其他物理学家的共同看法。比如著名美国物理学家奥本海默（J. Robert Oppenheimer, 1904—1967年）在为纪念爱因斯坦逝世10周年而撰写，后被收录于爱因斯坦诞辰100周年纪念文集《爱因斯坦——世纪文集》（*Einstein: A Centenary Volume*）的题为《论爱因斯坦》（*On Albert Einstein*）的文章中，就写过一段与英菲尔德的回忆有异曲同工之意的文字：

量子的发现必定会以这种或那种的方式出现……对没有任何信号能运动得比光更快的含义的深刻理解也必定会出现……直到今天仍未被实验很好证实的广义相对论则除他以外，在很长很长时间内都不会有人能提出。

《爱因斯坦——世纪文集》的主编、以编写物理教材而知名的美国麻省理工学院（MIT）物理学教授弗伦奇（Anthony French, 1920— ）也在为该文集撰写的题为《广义相对论的故事》（*The Story of General Relativity*）的文章中“英雄所见略同”地写道：

有人曾经评论过，在1905年之前……狭义相对论已经呼之欲出了，如果爱因斯坦没有将之透彻化，用不了多久其他人也会做到。无论这是否正确，可以确定的是：在创立广义相对论时爱因斯坦迈出了自己独一无二的一步。没有他的引路，这一步也许几十年都不会有人迈出。

这样一场“一个人的战斗”从历史考究的角度看，照说是不该有什么

悬疑的，其实却不然。在广义相对论的历史考究中除了探讨爱因斯坦的“心路历程”外，还有一个颇有争议性的话题，那便是究竟谁最先提出了广义相对论场方程？这个争议性话题就是本文的主题，它的主角有两位：一位当然是爱因斯坦，另一位则是德国数学家希尔伯特（David Hilbert，1862—1943年）。

二、希尔伯特对物理学的兴趣

希尔伯特是20世纪最著名的数学家之一，也是“数学圣地”哥廷根

（Göttingen）的灵魂人物之一，不仅研究领域极为宽广，研究成果也极为丰硕。单就以他名字命名的数学名词而论，就有不下一打，比如希尔伯特基（Hilbert basis）、希尔伯特特征函数（Hilbert's

characteristic function）、希尔伯特立方（Hilbert cube）、希尔伯特矩阵（Hilbert matrix）、希尔伯特模形式（Hilbert modular form）、希尔伯特函数

（Hilbert function）、希尔伯特多项

式（Hilbert polynomial）、希尔伯特概型（Hilbert scheme）、希尔伯特空间（Hilbert space）、希尔伯特变换（Hilbert transform）、希尔伯特不变积分（Hilbert invariant integral），等等，以及——最后但对本文来说绝非最不重要的——爱因斯坦-希尔伯特作用量（Einstein-Hilbert action）。



德国数学家希尔伯特

对于很多数学家来说，名字能出现在一个数学名词中就已是难得的荣誉了，但对希尔伯特来说，那一打以上的数学名词加在一起，也还只是勾勒出了他研究工作中偏于“战术性”的那部分，而未能包括很多视野更宏大的“战略性”研究——比如对几何基础及数学基础的研究。不仅如此，作为数学家的希尔伯特的研究领域甚至不是数学所能涵盖的，因为

除数学外，他对物理学也怀有浓厚的兴趣并从事过研究。

早在1900年发表的著名演讲“数学问题”（*Mathematische Probleme*）中，希尔伯特就把物理学的公理化列为了问题之一（即希尔伯特第六问题）。这个貌似泛泛的问题并非是为了让他的演讲看起来包罗万象而随意引入的，而确确实实是代表了希尔伯特所看重并感兴趣的一个方向。希尔伯特后来的学术轨迹在很大程度上印证了这一点：自1902年起，他开始讲授物理学；自1912年起，他设立了物理学助手职位，并招收指导了从事理论物理研究的学生；1913年，他组织了所谓的“哥廷根周”（*Göttinger Gastwoche*）活动，邀请普朗克（*Max Planck*, 1858—1947年），德拜（*Peter Debye*, 1884—1966年），能斯特（*Walther Nernst*, 1864—1941年），索末菲（*Arnold Sommerfeld*, 1868—1951年），洛伦兹（*Hendrik Lorentz*, 1853—1928年）等众多第一流的物理学家来做报告，介绍了包括气体运动理论及兴起中的量子论在内的诸多课题。1914年，他邀请物理学家德拜开设了有关物质结构的讲座。

在希尔伯特对物理学的兴趣中，公理化思想是一个很重要的方面，不仅他本人深为重视，受他影响，一些其他数学家也对物理学的公理化展开了研究。比如在哥廷根大学（*University of Göttingen*）就读过的希腊数学家卡拉西奥多里（*Constantin Carathéodory*, 1873—1950年）在热力学的公理化方面就做了重要工作。除公理化思想外，极小值原理（*minimal principle*）也极受希尔伯特的器重。极小值原理在物理学上的具体应用有着各种不同形式，那些形式大都为希尔伯特所熟悉。比如在讲授力学时，他曾经使用过高斯最小约束原理（*Gauss' principle of least constraint*）；在我们将要介绍的有关引力理论的研究中，则使用了在现代物理中被广泛运用的最小作用量原理（*principle of least action*）。

在对物理学的持续关注中，希尔伯特那颇具识人之明的眼光并没有

漏掉一位比他年轻17岁、正快速成长为大腕的“后起之秀”——爱因斯坦。早在1912年，希尔伯特在研究线性积分方程时，就曾与爱因斯坦有过信件往来：他向爱因斯坦索要过气体运动理论及辐射理论方面的论文，并回赠过一本自己新出版的积分方程著作。他也曾邀请爱因斯坦在“哥廷根周”期间访问哥廷根，做一次有关气体运动理论的报告，但爱因斯坦婉拒了。不过，在1915年6月28日至7月5日之间，爱因斯坦终于应希尔伯特的邀请对哥廷根进行了为期一周的访问，并作了六次——每次两小时的——报告，介绍他的广义相对论研究。



那次哥廷根之行给爱因斯坦留下了不错的印象，他在1915年7月15日给索末菲的信中描述了自己的观感：

在哥廷根，我非常愉快地看到所有的东西都在细节上得到了理解。我对希尔伯特很是着迷，他是一个重要人物！

在给其他同事和朋友的信件中，爱因斯坦也毫不讳言地表示了对希尔伯特的好感，并提到他（在引力理论方面）已完全说服了希尔伯特与克莱因（Felix Klein, 1849—1925年）⁽²⁾。希尔伯特对爱因斯坦的访问也极为重视。那次演讲之后不久，希尔伯特离开了哥廷根去度暑假。对于他在那段时间里的具体行程，史学家们所知不全，但一般认为，在那段时间里希尔伯特的研究重心向引力理论方向作了显著倾斜。这一倾斜使得他与爱因斯坦之间展开了一场无形的——在某些环节上甚至是有形的——竞争，也为史学家们留下了一个小小的谜团。

1915年11月20日，希尔伯特在哥廷根皇家科学院（Royal Academy of Science in Göttingen）作了有关引力理论的报告，介绍了他的研究成果。那次报告对于探讨谁最先提出了广义相对论场方程是极为重要的。可惜的是，也许因为听众大都是数学家，报告的主题却是物理学，从而“言者谆谆，听者藐藐”的缘故，后世的史学家们未能收集到有关那次报告的第一手资料——比如听众的反响或有关报告内容的细节性回忆等。早期的史学研究所依据的乃是希尔伯特于1916年3月31日发表在《皇家科学与人文学会新闻》（*Nachrichten von der Königlichen Gesellschaft der Wissenschaften*）上的题为《物理学基础》（*Die Grundlagen der Physik*）的论文。⁽³⁾那篇论文明确标注了曾在1915年11月20日的会议上作过报告（Vorgelegt in der Sitzung vom 20 November 1915），从而被视为了有关那次报告的最直接——一度甚至是唯一直接——的资料。

接下来我们就先对希尔伯特的那篇论文作一个简短介绍。

三、希尔伯特的《物理学基础》

希尔伯特对引力理论的研究有两个主要切入点：一个是来自爱因斯坦的广义协变原理及用度规张量描述引力势的想法，这是爱因斯坦自1907年开始思索引力问题以来逐步确立起来的想法；另一个则是来自德国物理学家米（Gustav Mie, 1869—1957年）的物质理论。米（一个字的中文译名真是别扭）的物质理论（简称米理论）是建立在物质起源于电磁相互作用这一被称为电磁观或电磁世界观（electromagnetic worldview）的观念之上的，⁽⁴⁾与建立在电磁观之上的其他理论一样，在昙花一现之后很快就入住了“历史博物馆”。

借助这两个切入点，沿袭深受其重视的物理学公理化的大思路，希尔伯特在论文的开篇中列出了两条公理。其中第一条被称为“米的世界函数公理”（Mie's axiom of the world function），这条公理虽被冠以米的名字，从框架上讲，实际引入的乃是最小作用量原理，只不过用被米称为“世界函数”的函数 H 来表述作用量而已。⁽⁵⁾除此之外，该公理还规定世界函数 H 只包含度规张量及其一、二阶导数，以及电磁势及其一阶导数。从而同时体现了爱因斯坦用度规张量描述引力势的想法以及米的建立在电磁观之上的物质理论（因为物质场部分只含电磁场）。而第二条公理则是所谓的“广义不变性公理”（axiom of general invariance），它规定世界函数在任意坐标变换之下为标量。毫无疑问，这条公理体现的是爱因斯坦的广义协变原理，只不过作用量是标量，从而“协变”（covariance）成为了“不变”（invariance）。如果更细致地分析的话，那么第一条公理中的 H 只包含度规张量及其一、二阶导数的限定有可能也是来自爱因斯坦的，因为他通过对经典极限的研究，发现了引力理论不含度规张量的二阶以上导数。⁽⁶⁾

以这两条公理为基础，希尔伯特给出了一系列数学和物理上的结果。其中数学上的结果包括：

（1）缩并形式——即关于里奇张量（Ricci tensor）——的毕安基恒等式（Bianchi identity）。[\(7\)](#)

（2）诺特定理（Noether's theorem）的雏形。[\(8\)](#)

而对我们来说更有兴趣的物理上的结果则主要包括：

（1）引力理论的作用量为 $K+L$ ，其中 K 为曲率标量（现代符号为 R ）， L 为物质场的作用量，对希尔伯特来说特指为米理论中的电磁作用量。

（2）引力场方程为 $\sqrt{g}(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}) = -\partial(\sqrt{g}L)/\partial g^{\mu\nu}$ ，其中 $K_{\mu\nu}$ 为里奇张量（现代符号为 $R_{\mu\nu}$ ）， $g_{\mu\nu}$ 为度规张量， g 为度规张量的行列式。

这两个物理上的结果正是后来在史学界引发争议乃至风波的核心所在。如前所述，包含这两个结果的论文虽是1916年3月31日发表的，但由于明确标注了曾在1915年11月20日的会议上作过报告，而关于那次报告，又一度并无足够详尽的其他资料可供研究，因此早期的史学家们便将这篇论文中的结果视为是希尔伯特不迟于1915年11月20日所得到的。与之相比，爱因斯坦最早得到正确的引力场方程是在1915年11月25日，那一天他向普鲁士科学院（Prussian Academy）报告了正确的场方程，并随即以“引力场方程”（*The Field Equations of Gravitation*）为题发表了普鲁士科学院的会议报告（*Sitzungsberichte*）中。这一时间比希尔伯特作报告的日子晚了5天。因此，一些早期的史学家认为希尔伯特先于

爱因斯坦就得到了广义相对论场方程。

不过，对于多数其他人来说，希尔伯特的论文其实并未引起太大反响，这也许是因为——如前所述——聆听他1915年11月20日报告的大都是数学家，对物理学话题相对隔膜。而当希尔伯特的论文正式发表时，不仅爱因斯坦关于引力场方程的最早的短文早已发表，就连他有关广义相对论的著名长篇综述《广义相对论基础》（*The Foundation of General Theory of Relativity*）也已问世。另外一个可能的原因则是爱因斯坦研究广义相对论所用的数学对当时的物理学家来说虽有些另类乃至“高深”，但比起希尔伯特的数学来却可能还算是略显“通俗”的，从而更容易被接受。

由于希尔伯特的论文未引起太大反响，因此关于希尔伯特是否先于爱因斯坦得到广义相对论场方程一事，很多人即便风闻过消息，对细节也大都知之不详。这方面的一个例子，是印度裔美国科学史学家梅拉（Jagdish Mehra, 1937— ）所提到的。1974年，梅拉在有关这一话题的著作《爱因斯坦、希尔伯特与引力理论》（*Einstein, Hilbert, and the Theory of Gravitation*）的序言中提到，他之所以研究这一课题，并撰写这一著作，是因为匈牙利裔美国数学及物理学家维格纳（Eugene Wigner, 1902—1995年）曾风闻过希尔伯特先于爱因斯坦发现广义相对论场方程的说法，并向他寻求证实。维格纳不仅是数学及物理学家，而且曾在哥廷根大学做过希尔伯特的助手，连他都不知道此事的细节，其他人就更可想而知了。

但在转而探讨历史细节之前，有关希尔伯特那篇论文还有其他一些东西值得评述。

在希尔伯特的论文中，除上述结果外，还提出了一个有趣的观点，

那就是引力场的作用量对于度规张量的10个分量和电磁势的4个分量分别作变分，一共可以得到14个方程（即10个引力场方程与4个电磁场方程），但由于缩并形式的毕安基恒等式共有4个，因此其中有4个方程是不独立的。希尔伯特对这一结果作出了自己的诠释，他认为这意味着4个电磁场方程可以作为10个引力场方程的推论，从而表明电磁理论可以从引力理论中得到。^[9]而依据希尔伯特当时所认同的电磁观，电磁理论乃是物质理论的基础，因此电磁理论可以从引力理论中得到，也就意味着全部的物理学都可以归并为引力理论。希尔伯特将自己那篇有关引力理论的论文取名为《物理学基础》，就在一定程度上体现了这一诠释。在论文的末尾，他甚至充满乐观地展望道：

通过本文所确立的基本方程式，我相信最深层的、目前还隐藏着的原子内部过程也将得到解释。尤其是将所有物理常数普遍约化为数学常数必定是可能的.....物理学在原则上变成像几何那样的科学——这毫无疑问是公理化方法的最高成就。

这种乐观憧憬也是刻在其墓碑上的希尔伯特的毕生信念“我们必须知道，我们必将知道（Wir müssen wissen. Wir werden wissen）”的一个生动写照。

当然，我们现在知道（希尔伯特本人在不久之后也意识到了），希尔伯特的上述看法是完全错误的，因为缩并形式的毕安基恒等式并不意味着电磁理论可以从引力理论中得到。从现代观点来看，4个缩并形式的毕安基恒等式的存在使得10个引力场方程中只有6个是独立的，从而在求解度规张量时必须添加4个坐标条件，仅此而已。这一点实际上是与广义协变性一脉相承的，因为后者意味着引力场方程及其解允许对4个时空坐标作任意变换，从而只有在添加4个坐标条件后才能得到确定的解。这一切丝毫不意味着从引力理论中可以得到电磁理论，更谈不上

能将全部的物理学归并为引力理论（后者还进一步假定了本身也是错误的电磁观），及支持希尔伯特论文末尾那些天马行空般的想象。此外，正确的引力场方程与缩并形式的毕安基恒等式一同确保了协变形式的能量动量守恒定律，而能量动量守恒定律——视具体的物质体系而定——蕴含了物质运动方程的全部或部分信息，这在如今也已是众所周知的结论了，前者在现代广义相对论教材中更是往往作为引力理论所需满足的条件，及推导引力场方程的捷径来用。可惜在早期研究中，无论爱因斯坦还是希尔伯特都未能清楚地看到这些。爱因斯坦早年走过的许多弯路（包括一度以为广义协变性无法普遍成立），以及希尔伯特的上述错误都与之不无关系。当然，这绝不能作为后人苛责他们的理由，在黑暗中探索的前辈们所面临的困难是我们这些事后诸葛无法直接体验的，爱因斯坦本人对此有过精辟的评论：

在黑暗中探寻真理的那些能够体味却难以描绘的年月，那些强烈的渴望和在信心与疑虑之间的反复徘徊，直至突破后的明晰和领悟，都只有亲身经历过的人才能知晓。

不过，希尔伯特将电磁理论乃至整个物理学归并为引力理论的观点虽然不正确，他的这一做法却可以算是先于爱因斯坦走上了试图统一引力与电磁的道路。当然，在这点上他虽先于爱因斯坦，却也绝非“第一人”，比如笃信电磁观的米在他之前就曾做过类似努力（参阅前注）。希尔伯特本人则将这条道路的开创归功于德国数学家黎曼（Bernhard Riemann，1826—1866年），表示黎曼是“最早探索引力与光之间的理论关联”的人（因为在黎曼手稿中有一篇探讨引力与光的短文），而他自己所得到的结果则被他称为是“对黎曼提出的问题的简单且很令人惊讶

的解答”。相比之下，爱因斯坦是20世纪20年代开始才正式走上同样道路的，不仅比黎曼、米、希尔伯特来得晚，也晚于德国数学家外尔（Hermann Weyl, 1885—1955年）、卡鲁查（Theodor Kaluza, 1885—1954年）等人——当然，他较晚进入这一“死胡同”对物理学来说乃是不幸中的幸运。

希尔伯特那篇论文的另一个值得评述的特点，是率先用最小作用量原理表述了正确的引力理论。希尔伯特以最小作用量原理为基本出发点（即视为公理）的做法，曾被奥地利物理学家泡利列为是妨碍物理学家接受他理论的两大障碍之一（另一个障碍是采用了米理论）。不过从现代物理学的观点来看，希尔伯特的做法却极具前瞻性。因为现代物理学上几乎所有的基础理论研究都是从最小作用量原理出发的。⁽¹⁰⁾就连爱因斯坦本人，虽然曾在1916年5月24日给好友艾伦菲斯特（Paul Ehrenfest, 1880—1933年）的信中表示不欣赏希尔伯特那“不必要地复杂迂回”的理论，在同年的10月却开始了沿这一方向的研究，并且在论文中改称希尔伯特的理论为“特别清晰的形式”。看来外尔在晚年的回忆中把希尔伯特比喻为吹着迷人长笛，引诱一大群老鼠跟随他跳入数学长河的人是颇为贴切的——就连爱因斯坦也挡不住诱惑地跟着他跳了一回。

说到这里，顺便回过头来评述一下本文开头所引述的爱因斯坦本人及其他物理学家的看法，即认为广义相对论如果没有爱因斯坦，在非常长的时间内都无法由别人提出。我早年接触到这一看法时，曾有过完全的认同，因为像广义相对论那样复杂的理论，是不可能像发现牛顿引力定律那样利用观测线索来发现的（那样的线索只能得到一部分后牛顿近似，而不可能反推出广义相对论来）；而另一方面，纯理论的探索又有太多的可能性，米等人的探索就是例子，甚至连爱因斯坦本人的探索，

也走过了大量歧途，且最后的成功——如后文将要提到的——仍有一定的歪打正着之处。对于这样的复杂理论，其他人完全独立地提出一模一样的理论似乎确实是不可思议的。⁽¹¹⁾不过，后来我的看法有了显著改变，理由正是最小作用量原理。从最小作用量原理的角度讲，只要有人想到了坐标变换可以突破狭义相对论的限制（这当然也不容易，但与创立整个广义相对论相比还是容易得多的），则度规张量的引入就是必然的，而度规张量及其低阶导数构成的最简单的标量就是曲率标量，这一数学事实也早晚会被注意到的。如果进一步考虑到在现代物理研究中，对最小作用量原理的运用越来越广泛，对作用量的选取则呈现出穷举性，即认为凡未被基本原理所禁止的项都可以进入作用量中，则曲率标量的进入——从而广义相对论的发现——也几乎是必然的。当然，历史只有一次，我的这种看法只能聊作谈资而已。

在结束对希尔伯特那篇论文的介绍之前，还有一点我愿评述一下，那就是一些人（包括爱因斯坦）在评论希尔伯特的理论时，往往会指出该理论不如爱因斯坦的理论来得普遍，理由是希尔伯特采信了建立在电磁观之上的米理论，从而使得物质场被局限于电磁场，丧失了普遍性。与之相比，爱因斯坦在1915年11月25日首次得到正确的场方程时，在论文中明确表示：“除了要求它导出守恒定律外，我们无需对物质场的能量张量作出其他假设。”这一评论所涉及的基本事实当然是无可辩驳的，希尔伯特确实对米理论表现出了显著青睐，甚至不惜以米的名字来命名公理。不过，我们也应当看到，将引力理论中的物质场限定为电磁场，甚至特指为米理论，更多地只是一种观念性的限制。希尔伯特的论文从理论框架上讲其实是相当普遍的，米理论虽然在公理中被提及，实质地位却是可有可无的，只要将作用量 $K+L$ 中的物质场部分 L 由仅仅描述米理论中的电磁场推广为一般的物质场，希尔伯特的理论框架无需任何修改就可以适用于更广泛的情形。不仅如此，希尔伯特的理论框架还

首次给出了能量动量张量的一个全新的表达式： $T_{\mu\nu} = (1/\sqrt{g})\partial(\sqrt{g}L)/\partial g^{\mu\nu}$ 。这个表达式用泡利的话说，是“只有在广义相对论中才变得显而易见的”。这个表达式自动保证了能量动量张量的对称性，从而有着独特的优越性。

当我们如今来回顾希尔伯特的理论时，出于尽可能精确地还原历史的目的，虽然毫无疑问地需要指出他曾经青睐过米理论，并将之列为公理这一事实，但却不应由此而忽视他的理论所具有的框架意义上的普适性。在历史上，理论的框架比创始者所设想的更为宏大的情形并不鲜见，比如1954年杨振宁（Chen Ning Yang, 1922— ）和米尔斯（Robert Mills, 1927—1999年）在提出著名的杨-米尔斯理论（Yang-Mills theory）时，就曾错误地将理论中的规范变换设定为同位旋（isospin）变换，但这种历史性的错误并不能遮蔽他们理论所具有的框架意义上的普适性，也并不妨碍我们将提出非阿贝尔规范理论（non-Abelian gauge theory）的荣誉归于他们。同样，对希尔伯特当年的论文，我们也不应该由于他所具有的观念上的局限性或错误，而减少应该归于他的荣誉——虽然希尔伯特的荣誉库早就“不差钱”了。

以上就是对希尔伯特那篇论文的已不太简短的“简短介绍”，现在回到我们的主题上来，即究竟谁最先提出了广义相对论场方程？或者更明确地说，希尔伯特是否先于爱因斯坦提出了广义相对论场方程？这可能是希尔伯特那篇遭受冷遇的论文中后世史学家们关心和讨论得最多的问题。这个问题除了关系到谁最先提出广义相对论场方程外，还可以引发一些其他的可能性，比如爱因斯坦在提出场方程的过程中，是否有可能受到过希尔伯特的启发？甚至是否有可能“借鉴”了希尔伯特的场方程（假如确实是后者先问世的话）？那样的可能性后来也确实被某些史学家提了出来，并引发了争议乃至风波——这些我们将在后文中提及。

四、早期研究简述

接下来，我们简单提一下在希尔伯特与广义相对论场方程这一课题上的早期研究。这一课题在早期虽不曾有过显著争论，但人们的看法起初也并不是完全一致的，比如洛伦兹就曾认为希尔伯特的工作只是用变分原理对爱因斯坦的工作做了重新表述而已。不过这些歧见在20世纪70年代出现了统一的势头。这势头的出现也许首先要归功于梅拉。如前所述，1974年，梅拉因受维格纳垂询而对这一课题进行了研究，并发表了《爱因斯坦、希尔伯特与引力理论》这一著作。不过，梅拉的研究虽缘起于维格纳对细节的垂询，视野却比较宏观，较少辨析历史细节，而更多地着眼于对广义相对论的历史，尤其是对希尔伯特所采用的思想方法及其来龙去脉进行整体阐述上。在梅拉之后，1978年，科学哲学家厄尔曼（John Earman, 1942— ）与格里莫尔（Clark Glymour）也涉猎了这一领域，发表了一篇题为《爱因斯坦与希尔伯特：广义相对论历史上的两个月》（*Einstein and Hilbert: Two Months in the History of General Relativity*）的文章。由于当时希尔伯特与爱因斯坦的通信已被公布，因此厄尔曼与格里莫尔的文章包含了一些辨析性的内容。不过，侧重点和风格虽各有所异，那两组研究的结论是大体相同的，那就是基本肯定了希尔伯特先于爱因斯坦得到了广义相对论场方程。另外，那两组研究也肯定了希尔伯特与爱因斯坦在提出场方程的过程中虽有过交流，基本工作仍是彼此独立的。1982年，荷兰裔美国物理学家派斯（Abraham Pais, 1918—2000年）在其颇具影响力的爱因斯坦传记《上帝是微妙的》（*Subtle is the Lord*）中也做出了大体相同的判断，即：“爱因斯坦是广义相对论物理理论的唯一提出者，基本方程式的发现则应同时归功于他和希尔伯特。”

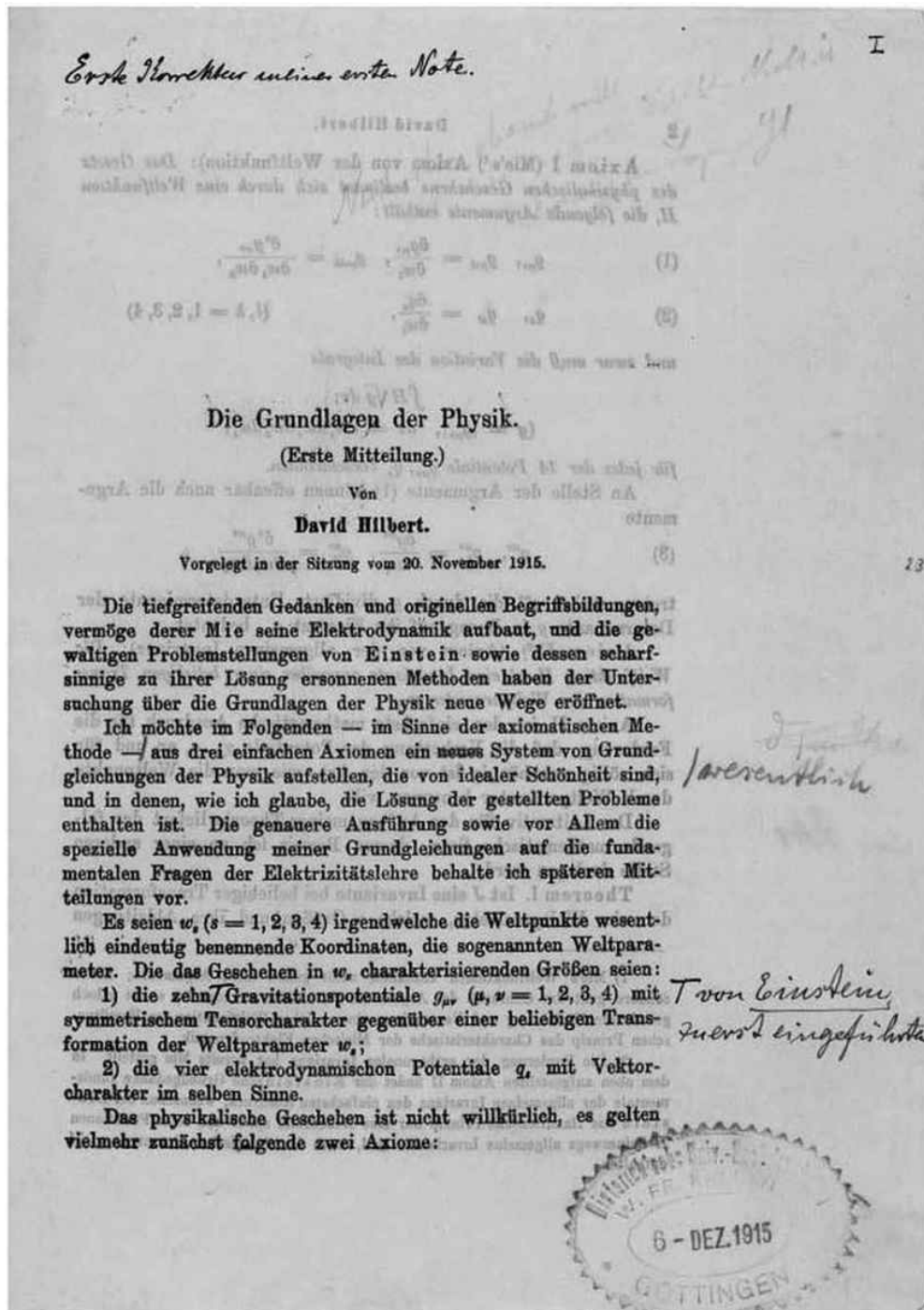
这些研究为早期的分歧作了“煞尾”。

但这种“煞尾”只维持了不太长的时间。1997年，以色列特拉维夫大学（Tel-Aviv University）的科里（Leo Corry）、德国普朗克科学史研究所（Max Planck Institute for the History of Science）的雷恩（Jürgen Renn）以及美国波士顿大学（Boston University）的施塔赫尔（John Stachel）一同在著名学术刊物《科学》（*Science*）上发表了一篇短文，标题为《希尔伯特-爱因斯坦优先权纠纷的迟到的裁决》（*Belated Decision in the Hilbert-Einstein Priority Dispute*）。这篇短文以一份从哥廷根档案馆得到的希尔伯特的论文校样为依据，将这一已尘封多年、堪称冷僻的陈年往事重新翻了开来，并引发了极大的争议。

五、校样风波

科里等人所得到的希尔伯特论文校样与发表稿一样，标注了曾在1915年11月20日的会议上作过报告。除此之外，它还带有一个打印日期：1915年12月6日，以及希尔伯特手注的说明“第一遍校样”（参阅图片），从而是一份比发表稿更早的文件。这一文件的发现，在一定程度上填补了研究这一课题所面临的早期文献空白。

通过对校样的研究，科里等人发现了希尔伯特在撰写发表稿过程中所做的若干修改，其中既包括逻辑结构的调整，比如校样包含三条公理，[\(12\)](#)而发表稿——如第三节所述——只含两条；也包括文字表述的修正，比如在首次提到表示引力势的度规张量时，手写加注了“由爱因斯坦最早引入的”这一说明。这些修改的存在，表明希尔伯特论文的发表稿与1915年11月20日的报告是有差异的，而早期研究将两者混为一谈的做法则是有缺陷的。



希尔伯特论文校样的第1页

问题是：缺陷大到什么程度？

对此，科里等人作出了一个他们称之为“迟到的裁决”的有一定颠覆

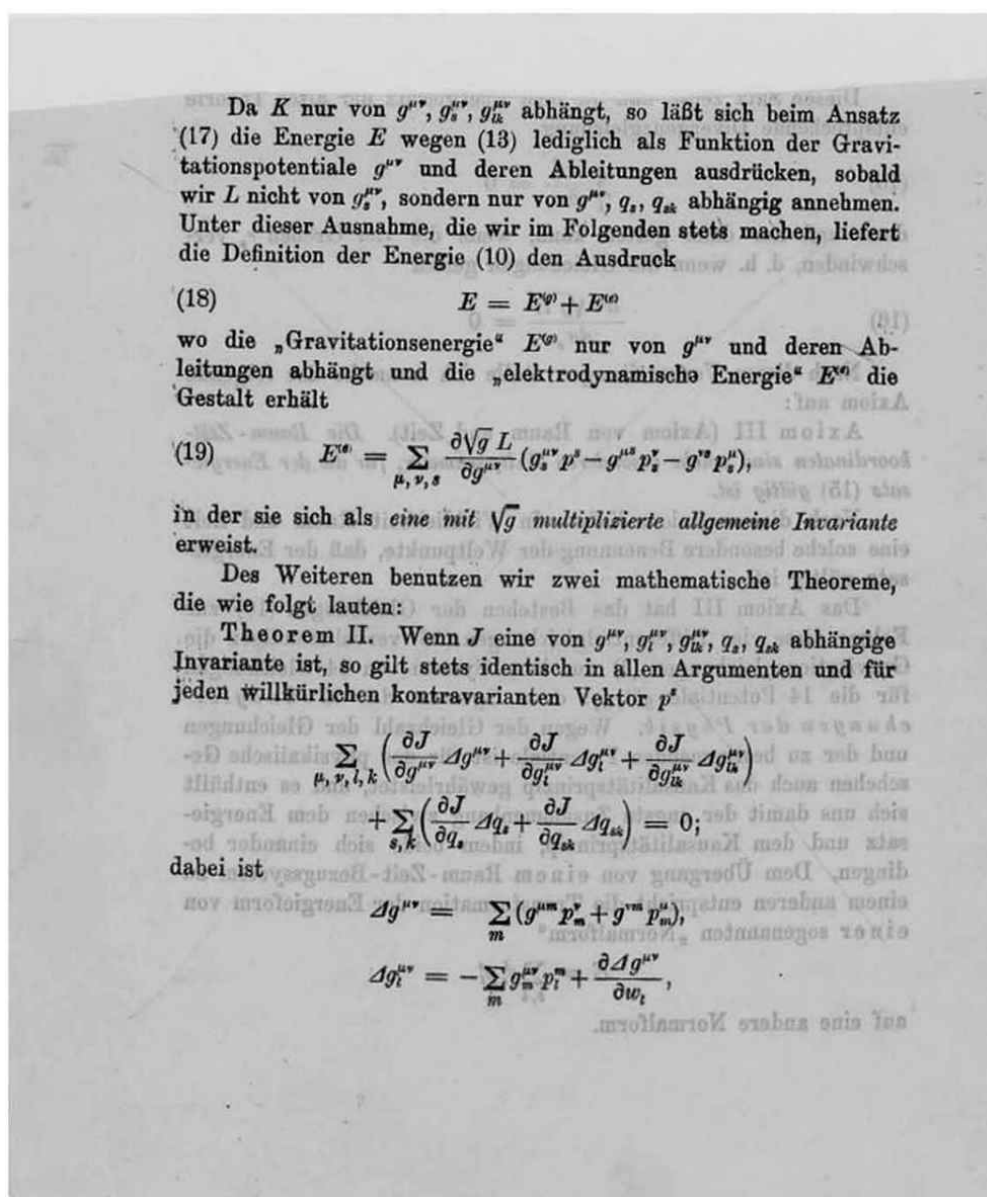
性的结论，那就是希尔伯特并未先于爱因斯坦提出广义相对论场方程。具体地说，科里等人发现在希尔伯特的论文校样中，只包含了我们在第三节中提到的两条物理结果中的第一条——即引力理论的作用量为 $K+L$ ，以及引力场方程可以通过对度规张量做变分而得到这一泛泛说明，却没有包含第二条——即引力场方程的具体形式。

以这一发现为基础，外加对另一处细节的分析（[参考此处注\(19\)](#)），科里等人还提出了一个更具颠覆性的观点，那就是在希尔伯特提出广义相对论场方程的过程中，有可能“借鉴”过爱因斯坦的场方程。这样，他们就不仅逆转了希尔伯特与爱因斯坦得到广义相对论场方程的时间顺序，还冲击了早期研究中一致认定的希尔伯特与爱因斯坦彼此独立地得到广义相对论场方程的结论，从而在史学界引起了极大的争议。

颠覆性结论引发极大的争议乃是常见现象，本不足为奇。不过，这场有关希尔伯特与广义相对论场方程的争议却有一个奇特之处，那就是它的切入点出现在了了一个几乎没人猜得到的地方。用旁观者的眼光来看，科里等人的研究中有一样东西堪称铁证，那就是希尔伯特的论文校样，这也是他们整篇文章的核心证据。但出人意料的是，争议的切入点居然就出现在了这核心证据上，并演变成了一场风波！

科里等人的短文发表后的第二年（1998年），加州理工学院（California Institute of Technology）的爱因斯坦文献专家索尔（Tilman Sauer）发表了一篇题为《发现的相对性：希尔伯特关于物理学基础的第一份笔记》（*The Relativity of Discovery: Hilbert's First Note on the Foundations of Physics*）的文章。在这篇文章中，他披露了一个令人吃惊的事实，那就是希尔伯特论文校样的第7、8两页（那是同一页的正反面）的上方有一部分是缺失的（参阅图片），而对引力场的作用量或场方程的叙述——如果有的话——恰好位于第8页的缺失部分中！

这情节简直赶上武侠小说了——在古龙小说《陆小凤传奇之金鹏王朝》中陆小凤怀疑有人冒充大金鹏王，判断的方法是查验脚趾，因为真的大金鹏王是六趾。但当陆小凤揭开了盖在大金鹏王腿上的毯子时，发现他的双腿是锯掉的！



希尔伯特论文校样的第8页

索尔虽披露了这一缺失，却未作发挥，只简单地假定为是希尔伯特

本人为了“偷懒”将之裁贴到了其他地方以节省时间。⁽¹³⁾但几年之后，这一缺失却不仅引发争议，而且演变成了风波。2003年，美国内华达大学（University of Nevada）的物理学教授温特伯格（Friedwardt Winterberg）发表了一篇题为《论〈希尔伯特-爱因斯坦优先权纠纷的迟到的裁决〉》的文章，对科里等人的短文进行了严厉驳斥，并对希尔伯特论文校样中的缺失作了阴谋论解读，认为缺失乃是近期发生的蓄意行为，目的是抹杀希尔伯特对广义相对论的贡献。这一阴谋论解读得到了另外几位作者的响应，比如美国作者毕尔克尼斯（Christopher Bjerknes）和德国作者沃恩茨（Daniela Wuensch）在分别于2003年和2005年出版的书中，对这一“阴谋”进行了详细剖析，甚至像福尔摩斯还原犯罪过程一样，“还原”了“阴谋”的各个环节，其中包括采用何种工具，以何种方式裁去文字，其间犯过何种错误，采取过何种补救措施，等等。他们甚至详细“还原”了“作案者”的心理活动。

这些阴谋论解读因细节丰富到了可笑的地步，就不在这里重复了。它们的致命伤不仅在于缺乏与细节相匹配的证据，更在于无法解释“作案动机”，即为何有人甘冒身败名裂的风险来对希尔伯特的论文校样做手脚？阴谋论者认为那是为了能提出如科里等人所提出的那种颠覆性的观点。但事实上，且不说在这一冷僻领域中提出颠覆性观点的收益与身败名裂的巨大风险根本就不成比例，即便真想提出颠覆性观点，也完全没必要做那样的手脚。因为在对缺失部分的篇幅及上下文的内容进行分析之后，史学界已提出了极强的分析理由，表明缺失部分只包含引力理论的作用量，而不包含引力场方程。这首先是因为引入后者及相关说明所需的额外篇幅绝非缺失部分所能容纳（发表稿中相应内容的篇幅也印证了这一点）；其次还因为在后文提到缺失部分中的公式时，表示由它对度规张量做变分可得到引力场方程，这说明缺失部分包含的是作用量，而不是场方程（发表稿同样印证了这一点，引力场方程是在作出上

述表示之后才给出的）。这一方面说明科里等人的核心证据虽一度成为争议的切入点，由铁证蜕变成了分析证据，实质内容却并无问题；另一方面也说明提出如科里等人所提出的那种颠覆性观点，根本就无需对校样作手脚。⁽¹⁴⁾此外，还有一点对阴谋论者很不利，那就是除无需对校样作手脚的科里等人外，这一领域并无其他人提出过类似的颠覆性观点（因此阴谋论的影射对象是极为鲜明的，那就是科里等人），从而根本找不到能从“阴谋”中获益的“作案者”，阴谋论的荒谬也就更不言而喻了。

另外有一点需要指出的是，阴谋论者在驳斥科里等人时，自己也提出了一个颠覆性观点，那就是明示或暗示爱因斯坦在提出广义相对论场方程的过程中，有可能“借鉴”过希尔伯特的场方程。⁽¹⁵⁾这一观点与科里等人的恰好相反，却同样冲击了早期研究中一致认定的希尔伯特与爱因斯坦彼此独立地得到广义相对论场方程的结论。这一对“互为镜像”的颠覆性观点，我们将在后文中加以讨论。

阴谋论提出后，被作为影射对象的科里等人在普朗克科学史研究所（即作者之一的雷恩所在的研究所）发表了一份针对温特伯格的措辞激烈的答复，不仅指责温特伯格的文字风格偏执（*paranoid*），而且将其观点与昔日纳粹德国的反相对论运动联系了起来，指控其试图重回纳粹时代反“犹太物理学”的老路。至此，争议演变成了风波。科里等人的这份答复因涉嫌人身攻击而遭到温特伯格的抗议，后来被普朗克科学史研究所“和谐”掉了，取而代之的是一份以研究所名义发表的对双方争议保持中立的声明。

由希尔伯特的论文校样引发的风波就介绍到这里，其基本结论是，希尔伯特的论文校样只包含引力理论的作用量，而不包含引力场方程。

六、信件辨析

由科里等人的短文引发的争议中，虽出现了荒谬的阴谋论和令人遗憾的人身攻击，却也涉及了一些细节性的辨析，且引起了不少认真讨论。接下来我们就对细节性辨析作一些介绍。

细节性辨析主要集中在对几封信件的解读上。其中第一封是1915年11月18日爱因斯坦给希尔伯特的信。这封信的背景是：希尔伯特于11月13日写信邀请爱因斯坦出席自己的报告，并表示该报告将对爱因斯坦提出的“大问题”（great problem）给出一个与爱因斯坦完全不同的公理化解答；爱因斯坦于11月15日回信以胃痛和疲惫为由婉拒；11月16日，希尔伯特给爱因斯坦发了一张如今已遗失了的明信片。爱因斯坦11月18日的信就是对该明信片的回复，其中包含这样的文字：

您给出的体系——就我所见——与我最近几个星期发现并向科学院报告过的完全一致。困难之处并不在于找到 $g_{\mu\nu}$ 所满足的广义协变方程，因为这可以在黎曼张量的帮助下很容易地得到，而是在于证明那些方程是一种推广，即对牛顿定律的简单而自然的推广。

针对希尔伯特表示自己的解答与爱因斯坦完全不同这一点，爱因斯坦这段文字有明显的宣示优先权并强调（自己的）物理重于（希尔伯特的）数学的意味。这段文字很早就引起过关注，但自科里等人用希尔伯特的论文校样推翻了原先基于发表稿作出的“希尔伯特先于爱因斯坦提出广义相对论场方程”的结论之后，这段文字有了更重要的地位，被一

些人视为了希尔伯特曾在11月16日的明信片上给出过广义相对论场方程的间接证据（因为否则的话，爱因斯坦所说的“完全一致”就失去了比较对象）。由于明信片比论文校样更早，假如它包含了场方程，那么希尔伯特就依然是先于爱因斯坦得到了广义相对论场方程。比如温特伯格就表示爱因斯坦的这段文字“证明了希尔伯特先于爱因斯坦得到了正确的场方程”；沃恩茨则以这段文字为由，用肯定的语气列举了希尔伯特明信片的内容，其中包括场方程。

但细究起来，事情却又不那么简单，因为爱因斯坦写下这段文字时他自己尚未得到正确的场方程。他当时以为正确的乃是一星期前（11月11日）发表的题为《关于广义相对论（补遗）》（*On the General Theory of Relativity (Addendum)*）的短文中提出的错误方程： $R_{\mu\nu} = -kT_{\mu\nu}$ ，它与正确方程之间相差一个正比于曲率标量的项： $-\frac{1}{2}g_{\mu\nu}R$ 。⁽¹⁶⁾这就产生了一个问题：既然爱因斯坦尚未得到正确的场方程，那么他所说的“完全一致”究竟是什么意思呢？对此，有人提出，是因为爱因斯坦当时正处于——用他自己的话说——“一生中最激动、最紧张的时期之一”，没时间仔细推敲希尔伯特的场方程，从而误以为它与自己的（错误的）场方程相一致。至于这种“误以为”的根源，有人认为是单纯的粗心大意，即漏看了希尔伯特的（正确的）场方程所多出来的一 $\frac{1}{2}g_{\mu\nu}R$ 项；也有人认为是爱因斯坦注意到了 $-\frac{1}{2}g_{\mu\nu}R$ 项在能量动量张量的迹 $T=0$ 时为零，而他当时恰好短暂地接受了能保证 $T=0$ 的电磁观（[参考此处注\(1\)](#)），从而认为该项的有无并不重要。⁽¹⁷⁾

这后一种猜测不无道理，在早期研究中也确实是对爱因斯坦所说的“完全一致”的合理解释。但希尔伯特论文校样的发现却为这一猜测带来了疑问，因为这一猜测假定了希尔伯特在11月16日的明信片上给出过场方程。如果那样的话，为什么在比明信片稍晚的论文校样中反而没有

场方程呢？对此，有人认为是场方程在希尔伯特心中没什么重要性。这种说法本身未必没有道理，希尔伯特有可能确实不像爱因斯坦那样重视场方程，但用它来说明这一疑问却是没有说服力的，因为希尔伯特若果真由于不重视场方程而未将之纳入论文校样，又怎会在小小的明信片上都不忘记写上呢？这些问题的存在，使得“明信片包含场方程”这一假定陷入了窘境。如果小结一下的话，那么这一假定的有利之处是可以解释爱因斯坦回信中的“完全一致”这一说法，不利之处则是无法解释论文校样不包含场方程。两者相比，不利之处显得更为突出，因为对“完全一致”这一说法可以有多种解释（下面就将给出另外一种），而且基于“明信片包含场方程”这一假定所做的解释虽不无道理，却也仅仅是不无道理而已；而论文校样不包含场方程这一点在“明信片包含场方程”的假定下却几乎不存在合理解释（尤其是从时间上讲，该明信片应该是对论文校样的概述，从而更不可能包含超乎后者的内容）。因此，“明信片包含场方程”这一假定虽被一些人寄予厚望，可靠性却并不高。

那么，希尔伯特的明信片究竟包含什么呢？从上面的分析来看，更有可能的只是引力理论的作用量，以及引力场方程可以通过对度规张量做变分而得到这一泛泛说明，而并不包含场方程。这不仅与论文校样相一致，而且也同样有可能解释“完全一致”这一说法。因为引力场的作用量对 $g^{\mu\nu}$ 的变分，若没时间推敲的话，是有可能被误认为或错算成 $R_{\mu\nu}$ ，从而与爱因斯坦的（错误的）场方程相一致的（事实上，除去无贡献的全微分项，作用量密度 R 的变分所给出的恰好是 $R_{\mu\nu}$ ，正确场方程中的额外项 $-\frac{1}{2}g_{\mu\nu}R$ 则是来自不变体积元中 \sqrt{g} 的变分，确实容易被忽视）。而且爱因斯坦当时甚至认为自己的（错误的）场方程是“唯一可能的广义协变方程”（他在11月18日的信中就强调了这一点），从而还有可能仅凭两者都广义协变这一特点就粗略地判断两者一致。当然，有人也许

会说：这种低级错误是爱因斯坦有可能犯的吗？对此当然谁也不敢肯定。高手虽然会犯低级错误（关于爱因斯坦犯低级错误的一个例子，[参考此处注\(16\)](#)），但没有铁证谁也不敢肯定某个特定的低级错误是高手犯的，而只能猜测，这是此类历史探究所无法避免的不确定性。但上述可能性比起爱因斯坦漏看希尔伯特场方程中的 $-\frac{1}{2}g_{\mu\nu}R$ 项来，恐怕还是大得多的可能性。另外，在我看来还有一个未被其他作者注意到的理由，可以支持希尔伯特的明信片不包含场方程这一结论，感兴趣的读者可以[参考此处注\(21\)](#)。

另一封引起广泛讨论的信件则是爱因斯坦1915年11月26日（即得到他自己的正确场方程的第二天）给好友仓格尔（Heinrich Zangger, 1874—1957年）的信，其中包含这样的文字：

这一理论的美丽是无可比拟的，然而只有一位同事真正理解了它，而他正以一种聪明的方式试图“分享”它（亚伯拉罕的表述）。在我的个人经历中，从未有比这一理论及相关一切所遭遇的更好地让我见识到了人性的卑劣。

这封信转译自《爱因斯坦全集》（*The Collected Papers of Albert Einstein*）第八卷的英译部分，其中“分享”一词来自英译partake（德文原词为nostrifizieren，另一种常用的英译为nostrify，含义为“吸纳”），对该词的引号来自原文。“亚伯拉罕的表述”是指德国大学将其他大学的学位“吸纳”为自己学位的做法，其中的“亚伯拉罕”是指经典电子论的代表人物、德国物理学家亚伯拉罕（Max Abraham, 1875—1922年）。“卑劣”一词则来自英译wretchedness。撇开可能因翻译而受损的精微词义不

论，关于这段文字，有两点是史学界公认的：一是“一位同事”指的是希尔伯特（对“一位”的着重来自原文）；二是这段文字有抱怨希尔伯特抢夺荣誉之意。除这封信外，爱因斯坦1915年11月30日给好友贝索

（Michael Besso, 1873—1955年）也写了一封信，其中提到“我的同事在此事中表现得很丑恶”（“丑恶”一词来自英译hideously），也印证了上述抱怨。

这些信也被一些认为爱因斯坦“借鉴”过希尔伯特场方程的人解读为希尔伯特先于爱因斯坦得到广义相对论场方程的证据，比如德国康斯坦茨大学（University of Konstanz）的物理学教授艾伯纳（Dieter W. Ebner）在2006年发表的文章《希尔伯特如何先于爱因斯坦发现爱因斯坦方程及对希尔伯特校样的伪造》（*How Hilbert has found the Einstein equations before Einstein and forgeries of Hilbert's page proofs*）中就这样解释爱因斯坦的愤怒：“他必定对自己为解决‘大问题’工作了8年，而希尔伯特只用几个星期就先于他优雅地找到答案感到非常愤怒。爱因斯坦的愤怒表明希尔伯特11月16日的明信片对他有过显著帮助。”这种解释的逻辑是：倘若希尔伯特没有先于爱因斯坦得到广义相对论场方程，爱因斯坦就不会如此愤怒，因此爱因斯坦的愤怒反证出了希尔伯特先于他得到广义相对论场方程。

这种解读的最大问题，除采信了希尔伯特明信片包含广义相对论场方程这一本身就并不可靠的假定外，还与世所公认的爱因斯坦人品有着巨大矛盾。因为假如这种解读成立，那爱因斯坦就不仅“借鉴”了希尔伯特的场方程，而且还在给朋友的信中倒打一耙，反诬希尔伯特“丑恶”，其人品可就不是一般的无耻了。在没有铁证之前，对世所公认的爱因斯坦人品做出如此颠覆性的推测，是极度草率的。因此，这种以爱因斯坦的抱怨信为依据反证希尔伯特先于爱因斯坦得到广义相对论场方程的做

法，是非常可疑的。相反，如果我们认为世所公认的爱因斯坦的人品更有公信力的话，倒是恰恰应该由抱怨信推证出爱因斯坦不可能“借鉴”希尔伯特的场方程。

而更平和的说法也许是，无需对爱因斯坦的人品作任何极端假设，就有很多理由可以解释爱因斯坦为什么会写下那样的抱怨信。比如前面提到的误以为希尔伯特的体系与他自己的“完全一致”（如前所述，这并不意味着希尔伯特在明信片给出了场方程），就有可能对处于冲刺阶段的爱因斯坦造成极大的危机感和恼怒感，因为那意味着希尔伯特这位听了他几次报告后半路杀出的竞争者有可能分享优先权。某些传记作者喜欢渲染爱因斯坦对名利的超脱，其实爱因斯坦起码对广义相对论的优先权是非常在乎的，别说希尔伯特，就连对昔日的“亲密战友”格罗斯曼（**Marcel Grossmann**, 1878—1936年）的贡献，他也曾在1915年7月15日给索末菲的信中近乎冷酷地表示：“格罗斯曼永远无法宣称是（广义相对论的）共同发现者，他只是在数学文献方面引导过我，而对结果没有任何重要贡献。”⁽¹⁸⁾对格罗斯曼尚且如此，希尔伯特这样一个“陌生人”分享优先权的可能性，自然不会让爱因斯坦开心，在给朋友的信中有所流露也就不足为奇。他信中的措辞虽然严苛，但与对格罗斯曼的评价相比，考虑到彼此关系的亲疏之别，并不能算出格。而且他不仅隐去了希尔伯特的名字，所选择的倾诉对象也都是圈外人士（仓格尔是法医学教授，贝索是专利局职员）。更何况，他当时恐怕不会想到自己在几年之后会成为每封私信都被人细细推敲的公众人物。将这些因素综合起来，那两封抱怨信并没有什么不可理解的地方，以致于非得需要靠颠覆爱因斯坦的人品，及采信希尔伯特明信片包含广义相对论场方程那样的假定才能解释。

以上就是对受到较多关注的爱因斯坦信件的分析，其基本结论是：

爱因斯坦的信件并不构成对“希尔伯特先于爱因斯坦提出广义相对论场方程”的有效支持。

七、“借鉴”之争

在前文中，我们多次触及了希尔伯特与爱因斯坦是否彼此独立地得到广义相对论场方程的问题。这个问题在早期研究中曾有过肯定答案，但在科里等人的短文发表后却出现了两种针锋相对的否定看法：一种是认为希尔伯特有可能“借鉴”了爱因斯坦的场方程；另一种则认为爱因斯坦有可能“借鉴”了希尔伯特的场方程。下面我们就对这两种观点做一些介绍与分析。

认为希尔伯特有可能“借鉴”了爱因斯坦场方程的观点从某种意义上讲，是早期就有过的诸如“希尔伯特的工作只是用变分原理对爱因斯坦的工作作了重新表述”（参阅第四节）之类泛泛看法的延续。只不过利用希尔伯特的论文校样不包含引力场方程这一新证据，具体提出了希尔伯特在看到爱因斯坦11月25日的论文之后才添加场方程，且在添加过程中“借鉴”了爱因斯坦场方程的观点。这种观点的前半部分从现有资料来看是成立的，因为希尔伯特的论文确实是在校样之后才添加场方程的，不仅时间上晚于爱因斯坦11月25日的论文，而且还援引了后者，从而表明希尔伯特看过爱因斯坦11月25日的论文（以希尔伯特当时对引力理论的关注，这是显而易见的）。在这种情况下，希尔伯特确实有可能因受到爱因斯坦11月25日论文的影响，而对自己的论文作出调整。不过，认为希尔伯特在添加场方程的过程中“借鉴”了爱因斯坦的场方程，却仍是缺乏依据的。

在说明这一点之前，让我们稍稍离题一下，先介绍另一个本身也值得一提的争议。科里等人的短文发表之后，有人提出了这样一个观点，那就是希尔伯特既然给出了广义相对论的作用量，就应该算是给出了场方程，因为后者不过是前者的变分而已，而且那变分用某些持这一见解

的人的话说，乃是“普通研究生就能完成的常规练习”。这个说法有道理吗？应该说既有道理又没道理。说它有道理，是因为作用量确实可以算是间接确定了场方程，而变分计算也确实不是很困难的问题；说它没道理，则是因为如今很简单的东西在初次探索时未必简单。比如广义相对论的牛顿极限如今正是“普通研究生就能完成的常规练习”，当年却让爱因斯坦和格罗斯曼栽了跟斗（[参考此处注\(16\)](#)）。广义相对论中的变分计算也是如此，起码在当年绝非是“普通研究生就能完成的常规练习”（[参考此处注\(21\)](#)）。另外，对广义相对论来说，场方程是一切物理计算的基础，重要性远高于作用量，这一点与现代读者所熟悉的量子场论之类以作用量为核心的理论完全不同。因此，在广义相对论的场方程尚未被提出之时，将它的提出归附于作用量的提出是很不恰当的。

现在回到希尔伯特对场方程的推导上来，这一推导实质上就是计算变分。如上所述，这在当年绝非是“普通研究生就能完成的常规练习”。但尽管如此，要说像希尔伯特那样的数学大师在决定由变分原理推导场方程时，会要“借鉴”爱因斯坦的场方程，却是令人难以置信的。而且，爱因斯坦在短短一个月的时间里就提出了好几种不同的场方程，我们这些事后诸葛虽然知道11月25日那个是正确的，对当年的希尔伯特来说却未必显而易见，他又怎会将自己的荣誉压在上面，“借鉴”其结果而不亲自计算变分呢？不仅如此，爱因斯坦对正确场方程的推导——如第三节所提到，并即将介绍的——有一定的“歪打正着”的意味，与希尔伯特对严密性的要求相距甚远，极不可能使他萌生“借鉴”之意。因此，希尔伯特得到广义相对论场方程的时间虽晚于爱因斯坦，说他“借鉴”爱因斯坦的场方程却是缺乏依据的。[\(19\)](#)

认为爱因斯坦有可能“借鉴”希尔伯特场方程这一观点则基本上是近年才出现的。在早期，哪怕是曾经显著贬低过爱因斯坦狭义相对论贡献

的英国数学家惠特克（Edmund Whittaker, 1873—1956年），在其1953年出版的《以太和电学的历史》（*A History of the Theories of Aether and Electricity*）的第二卷中述及广义相对论时，也不曾提出过那样的观点。持这一观点的人除了利用我们前面辨析过的那些信件外，还提出了一个理由，那就是爱因斯坦给出正确的场方程时没有进行推导。前文提到过的艾伯纳、哥廷根大学的理论物理学家托德洛夫（Ivan T. Todorov）等人都持此见。但这种理由却是相当粗心的，因为事实上，爱因斯坦在给出正确的场方程时是进行了推导的，只不过那推导有一定的歪打正着之处。

粗略地说，爱因斯坦的推导是这样的：利用 $\sqrt{-g}=1$ 这一他在那些年时常使用的简化条件，将真空场方程 $R_{\mu\nu}=0$ 用一个被他称为引力场的“能量分量”（energy component）的量 $t_{\mu\nu}$ 来改写，在改写的结果中出现了 $t_{\mu\nu} - \frac{1}{2}g_{\mu\nu}t$ 这一组合。然后他假定物质场的能量动量张量 $T_{\mu\nu}$ 与引力场的“能量分量” $t_{\mu\nu}$ 应该以相同方式出现在理论中。这就导致了 $T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T$ 这一构成正确场方程 $R_{\mu\nu} = -k\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right)$ 的关键组合。⁽²⁰⁾之所以说这一推导有一定的歪打正着之处，是因为引力场的“能量分量” $t_{\mu\nu}$ 并非张量，而物质场的能量动量张量 $T_{\mu\nu}$ 则是张量，假定两者以相同方式进入理论是没有充分依据的。爱因斯坦自己也在1916年1月17日给洛伦兹的信中承认：“基本表述终于正确了，但推导仍是糟糕的。”⁽²¹⁾确实，爱因斯坦直至1916年11月25日为止的所有广义相对论论文在结构上都远不如希尔伯特的论文严整。但正是那种不严整，与他过去若干年中近乎于用试错法寻找场方程的努力一脉相承，从而深具“爱因斯坦特色”（这一点爱因斯坦自己也知道，并在给洛伦兹的那封信中表示“我那一系列有关引力的论文是一条错误足迹的链条”）。更何况，

爱因斯坦“借鉴”希尔伯特场方程的前提是“希尔伯特先于爱因斯坦得到广义相对论场方程”，而如前所述，目前的证据并不支持这一前提。因此，认为爱因斯坦有可能“借鉴”希尔伯特的场方程也同样是缺乏依据的。

以上是对希尔伯特与爱因斯坦在推导场方程这一环节上是否相互“借鉴”的分析，其基本结论是：希尔伯特与爱因斯坦在推导广义相对论场方程这一环节上是相互独立的。

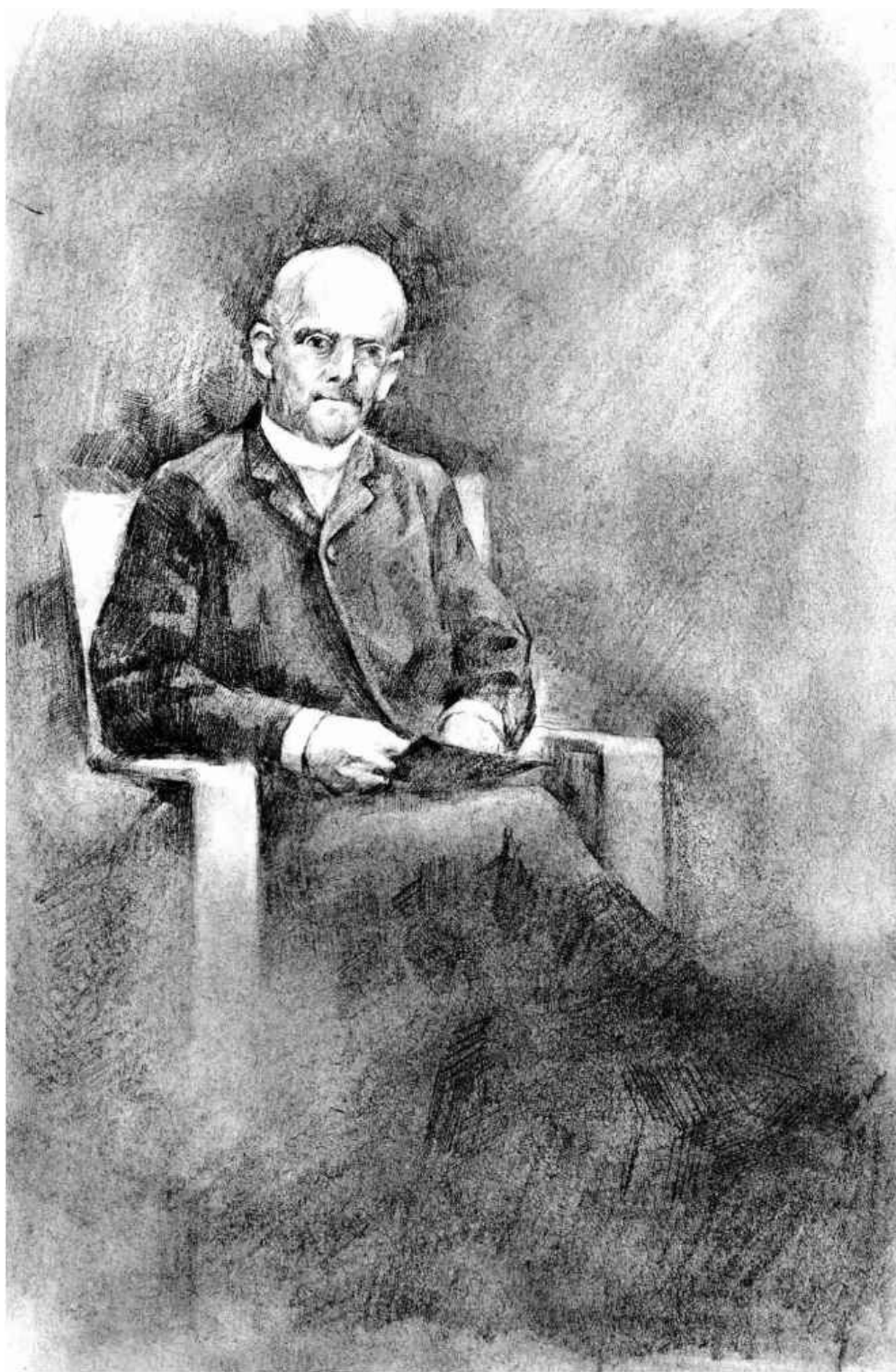
八、尾声

至此，本文就接近尾声了。由希尔伯特论文校样激起的争议也许还不能算是尘埃落定，但倘若没有新的原始资料被发现，那么透过飞扬的尘土我们已可大致窥见结果。总体来说，与早期研究相比，以较高可信度被改变的结论只有一个，那就是希尔伯特并未先于爱因斯坦得到广义相对论场方程。早期研究中的其他结论，比如希尔伯特与爱因斯坦在推导广义相对论场方程这一环节上是彼此独立的，则虽有新争议，却不足以推翻早期结论；而“爱因斯坦是广义相对论物理理论的唯一提出者”（派斯语）则更是无可撼动。

如果考虑到一些人提出的“给出了广义相对论的作用量，就应该算是给出了场方程”这一并未完全没有道理的观点，对那段历史中的优先权部分也许可以作这样一个无争议的表述：希尔伯特最先得到了广义相对论的作用量。有意思的是，对这段历史的研究虽不无波折，学术界对若干术语的命名却似乎早就预见到了结果：广义相对论的场方程通常被称为爱因斯坦场方程（Einstein field equations），而作用量则被称为爱因斯坦-希尔伯特（或希尔伯特-爱因斯坦）作用量（最先得到作用量的虽是希尔伯特，爱因斯坦作为整个理论的奠基者，享有共同“冠名权”应该不算过分）。

在本文的最后，让我们提一下那段历史对希尔伯特与爱因斯坦个人关系的影响，爱因斯坦给仓格尔和贝索的抱怨信可能会让某些读者为这两位大师级人物的关系捏一把汗，担心他们之间会出现“李杨之争”那样的局面。不过幸运的是，爱因斯坦的抱怨信虽折射出对优先权的敏感，他的整体人品与智慧终究还是让他摆脱了优先权之争的泥潭。而更值得称道的则是希尔伯特，他不仅在诸多场合公开承认广义相对论是爱因斯

坦的理论，而且还在评选第三届波尔约奖（Bolyai Prize）时推荐了爱因斯坦，并提名使爱因斯坦当选为了哥廷根数学学会（Göttingen Mathematical Society）的通讯会员（corresponding member）。[\(22\)](#)1915年12月18日，希尔伯特写信将当选消息告知了爱因斯坦。两天后，爱因斯坦给希尔伯特回了这样一封信：[\(23\)](#)



感谢您友好地告知我当选了通讯会员。借此机会，我觉

得有必要跟您说一件对我来说比这更重要的事情。我们之间近来有着某种我不愿分析其原因的不良感觉。我一直在努力抵御这种感觉带来的苦涩，并取得了完全的成功。我已在心中恢复了与您的往日友谊，并希望您也这样待我。两个已在一定程度上将自己从这个肮脏世界中解脱出来的真正的研究者若不能彼此欣赏，那将是一种真正的耻辱。

让史学界争议了近一个世纪的话题在那两位睿智的当事人之间，就这样放下了。

他们都放下了，我们还放不下吗？就让这封信也成为本文的终结吧。

参考文献

[1] L Corry, et al. Belated decision in the Hilbert-Einstein priority dispute [J] . Science, 1997, 278: 1270.

[2] Earman J, Glymour C. Einstein and Hilbert: two months in the history of general relativity [J] . Arch. Hist. Exact Sci., 1978, 19: 291.

[3] Ebner D W. How Hilbert has found the Einstein equations before Einstein and forgeries of Hilbert's page proofs [J/OL] . <http://arXiv.org/arXiv:physics/0610154v1> [physics.gen-ph] .

[4] Einstein A. 爱因斯坦文集（第一卷） [M] . 北京：商务印书馆，2009.

[5] Einstein A. 爱因斯坦全集（第五卷） [M] . 长沙：湖南科学技术出版社，2002.

[6] Einstein A. The Collected Papers of Albert Einstein (Vol 6) [M] . Princeton: Princeton University Press, 1997.

[7] Einstein A. The Collected Papers of Albert Einstein (Vol 8) [M] . Princeton: Princeton University Press, 1998.

[8] French A P. Einstein: A Centenary Volume [M] . Cambridge: Harvard University Press, 1980.

[9] Infeld L. Albert Einstein: His Work and Its Influence on Our

World [M] . New York: Charles Scribner's Sons, 1950.

[10] Isaacson W. Einstein: His Life and Universe [M] . New York: Simon & Schuster, 2008.

[11] Kox A J. The Universe of General Relativity [M] . Berlin: Birkhäuser, 2005.

[12] Logunov A A. et al. How were the Hilbert-Einstein equations discovered? [J/OL] . <http://arXiv.org/arXiv:physics/0405075v3> [physics.hist-ph] .

[13] Mehra J. Einstein, Hilbert, and the Theory of Gravitation: Historical Origins of General Relativity Theory [M] . Dordrecht: D. Reidel Publishing Company, 1974.

[14] Pais A. Subtle is the Lord [M] . Oxford: Oxford University Press, 1982.

[15] Pauli W. Theory of Relativity [M] . Oxford: Pergamon Press Ltd., 1958.

[16] Petrov V A. Einstein, Hilbert and equations of gravitation [J/OL] . <http://arXiv.org/arXiv:gr-qc/0507136v1>.

[17] Reid C. Hilbert-Courant [M] . New York: Springer-Verlag New York Inc., 1976.

[18] Renn J, Schemmel M. The Genesis of General Relativity (Vol 4) [M] . Berlin: Springer, 2007.

[19] Sauer T. The relativity of discovery: Hilbert's first note on the foundations of physics [J/OL] . <http://arXiv.org/arXiv:physics/9811050v1> [physics.hist-ph] .

[20] Thorne K S. Black Holes and Time Warps: Einstein's Outrageous Legacy [M] . New York: W. W. Norton & Company, 1994.

[21] Todorov I T. Einstein and Hilbert: the creation of general relativity [J/OL] . <http://arXiv.org/arXiv:physics/0504179> [physics.hist-ph] .

[22] Whittaker E. A History of the Theories of Aether and Electricity: The Modern Theories 1900—1926 [M] . Nashrille: Thomas Nelson and Sons Ltd., 1953.

[23] Winterberg F. On “belated decision in the Hilbert-Einstein priority dispute” [J] . Z. Naturforsch., 2004, 59a: 715.

二零一二年八月十五日写于纽约

(1) 本文曾发表于《数学文化》2013年2月刊（山东大学与香港浸会大学合办）。

(2) 克莱因是一位比希尔伯特更资深的哥廷根的数学大师（不要与有相同中文译名的美国数学科普作家Morris Kline相混淆，此处及后文提到克莱因所指的是Felix Klein）。关于爱因斯坦认为自己说服了希尔伯特一事，信件中并未提及细节，如果用后来的情形来印证的话，那么希尔伯特主要是接受了爱因斯坦的广义协变原理及用度规张量描述引力势的想法，但对爱因斯坦理论的若干其他细节，尤其是当时尚不正确的场方程，是并不认同的。

(3) 细心的读者也许注意到了，这一论文标题与他的名著《几何基础》（*Grundlagen der Geometrie*）的标题是极为相似的。

(4) 希尔伯特早年曾倾向于接受所谓的机械观，后来才转而接受电磁观。自1913年之后，则

更具体地青睐于建立在电磁观之上的米理论。

(5) 米的“世界函数”是一种拉格朗日函数（Lagrangian），后者的符号通常取为 L 。拉格朗日函数在具体使用时分“拉格朗日量”（Lagrangian）与“拉格朗日密度”（Lagrangian density）两种主要情形，前者常用于离散体系，后者常用于场论，希尔伯特所用的是后者。拉格朗日密度有时也称为作用量密度（action density），其对不变体积元 $\sqrt{g}d^4x$ 的积分则称为作用量。不过为简洁起见，在不会混淆的情形下，我们将把两者统称为作用量。另外，依据约定的不同， \sqrt{g} 有时——比如在爱因斯坦的论文中——会被表示成 $\sqrt{-g}$ ，将不再说明。

(6) 不过场论中通常就不含场量的二阶以上导数，因此这一限定虽有可能来自爱因斯坦，但即便没有爱因斯坦，估计希尔伯特也会做出同样的限定。

(7) 毕安基恒等式据说最早乃是意大利数学家里奇（Gregorio Ricci-Curbastro, 1853—1925年）于1880年发现的，后来于1902年由意大利数学家毕安基（Luigi Bianchi, 1856—1928年）重新发现，并因此得名。在广义相对论的早期研究中，无论爱因斯坦还是希尔伯特都不知道这一恒等式，从而未能直接利用它来简化场方程的推导。

(8) 诺特定理是德国数学家诺特（Emmy Noether, 1882—1935年）于1918年发表的，不过其证明据说在1915年就完成了。希尔伯特在准备1915年12月20日的报告期间，曾经让诺特做他的助手，因此两人在诺特定理这一课题上可能有过讨论。

(9) 有意思的是，希尔伯特的这一结论实际上是逆转了米的观点。米是电磁观的贯彻者，他的理论是单纯的电磁理论，并一度希望将引力也归因于电磁理论（因此米也是统一引力与电磁的早期尝试者），这与希尔伯特所得到的结论恰好相反。米后来放弃了将引力归因于电磁理论的努力，转而尝试用四维矢量来描述引力势。

(10) 需要提到的是，爱因斯坦、洛伦兹等人在相对论研究中也使用过最小作用量原理，只不过没有像希尔伯特那样将其地位放得如此之高，而且也没能给出正确的引力理论作用量。

(11) 有读者可能会觉得希尔伯特就是一个反例。其实不然，因为——如正文所述——希尔伯特的研究是建立在来自爱因斯坦的许多框架性观点之上的，从而至多只是在我们将要讨论的推导场方程这一环节上独立于爱因斯坦，而并非在整个引力理论的研究上都独立。

(12) 希尔伯特论文校样所包含的第三条公理被称为“时空公理”（axiom of space and time），它将时空坐标限定为使能量守恒定律成立的坐标。科里等人认为这一公理的使用表明希尔伯特1915年11月20日的报告曾受到爱因斯坦早期观点的影响，认为广义协变性无法普遍成立，从而需要对时空坐标附加额外限定。这一解读后来受到了其他人的质疑，因与本文主题关系不大，就不详述了。希尔伯特本人在给克莱因的信中表示，在发表稿中舍弃这一公理乃是因为它“不够成熟”。

(13) 这一说法并非全无依据，因为希尔伯特曾经做过类似事情，即从一篇文稿中裁去一部分内容，贴到其他地方以节省时间。

(14) 不过，科里等人的短文只字不提校样中的缺失，依然是一件奇怪的事情。有些史学家猜测科里等人撰写论文时校样仍是完整的。但毕尔克尼斯声称自己曾写信向作者之一的施塔赫尔求证过，后者表示之所以未提校样中的缺失，乃是因为那篇短文只是一份不完整的初步报告（施塔赫尔等人在2002年确实发表过更详细的文章，其中提到了校样中的缺失）。这一信息如果可靠，则表明科里等人撰写短文时校样中的缺失已经存在，从而他们以陈述事实的语气来陈述校样中不含引力场方程这一需经分析才能得到的结论就显得有些难以理解。施塔赫尔的回答，即那篇短文只是一份不完整的初步报告，也并不能令人满意（比如毕尔克尼斯就指出该回答与短文标题中的“迟到的裁决”那样的强势用语不相吻合）。但奇怪归奇怪，语气上的问题毕竟不是铁证，这件事无论如何上升不到阴谋论的高度（因为科里等人的短文根本无需“阴谋”来作后盾）。

(15) 阴谋论者中的有些人原本就对爱因斯坦持有很负面的看法。比如毕尔克尼斯早在2002年就出版过一本书，书名为《爱因斯坦：无可救药的剽窃者》（*Albert Einstein: The Incurable Plagiarist*）。说到毕尔克尼斯，还有必要提到这样一点，那就是此人的某些观点有明显的妄想家色彩，比如认为人类正受到共济会（Freemasonry）等组织的威胁，而他自己的知识是拯救人类所必需的。

(16) 不难证明， $-\frac{1}{2}g_{\mu\nu}R$ 在能量动量张量的迹 $T=0$ 时为零，因此爱因斯坦缺了这一项的场方程在一般情形下虽是错误的，在 $T=0$ 时却成立。这一点爱因斯坦自己也有所察觉，因为他发现自己的场方程在 $T\neq 0$ 时会出现问题（无法满足能量动量守恒）。为避免这一问题，他破天荒地短暂接受了电磁观（因为电磁场的能量动量张量满足 $T=0$ ，而电磁观意味着物质起源于电磁，从而保证了物质场的能量动量张量也满足 $T=0$ ）。另外要指出的是，这一错误场方程其实三年前就由爱因斯坦与格罗斯曼（Marcel Grossmann, 1878—1936年）一同提出过，只是因为计算牛顿极限时出现错误而遭到了放弃（这是爱因斯坦犯低级错误的一个例子）。爱因斯坦在11月18日的信中特意强调了这一点（目的也许是巩固优先权），从而证实了他信中所指的自己的场方程乃是这一错误方程。

(17) 从逻辑上讲还有一种可能性，那就是希尔伯特在明信片中给出了一个跟爱因斯坦错得一模一样的场方程。不过，由于希尔伯特的任何文稿中都从未出现过那样的错误方程，其理论体系中也看不到出现这种错误的温床，忽略这一可能性是有较充足的理由的。

(18) 爱因斯坦在晚年时对格罗斯曼表示出了更多的感激。比如1936年格罗斯曼去世后，爱因斯坦在给格罗斯曼遗孀的信中表示“感谢他（格罗斯曼）和他父亲的帮助，……要不然，即使未必死去，我也会在智力上被摧毁了。”而在爱因斯坦自己去世前一个月所撰的回忆中，他则写道：“我需要在自己在世时至少再有一次机会来表达我对格罗斯曼的感激之情。”不过，即使在这些表示中，他也没有在对广义相对论的贡献上给予格罗斯曼更高的评价。

(19) 这里还有一个细节需要提一下，那就是希尔伯特论文的发表稿在给出场方程时写了这样一句话（已改用现代记号）：“利用在由 $g^{\mu\nu}$ 及其一、二阶微商构成的量中， $R_{\mu\nu}$ 是除 $g_{\mu\nu}$ 之外唯一的二阶张量，以及 R 是唯一的不变量这一事实，无需计算就可得到。”这句话被科里等人视为是希尔伯特没有亲自计算变分，从而有可能“借鉴”爱因斯坦场方程的证据。应该说，这一看法不无道理，因为仅凭这句话是得不到场方程的，还必须辅以能量守恒的要求。但以此认定希尔伯特“借鉴”了爱因斯坦的场方程，比起正文所述的相反理由来，是仍显薄弱的，因为这有可能只是希尔伯特采取的一种在他看来更容易传递给读者的表述方式。很难想象在这种严肃论文中，崇尚严密性的希尔伯特会不进行严密的计算，而“借鉴”爱因斯坦那屡经变更且推导很不严密的场方程。

(20) 广义相对论场方程有两种常用的等价形式，爱因斯坦当时得到的是 $R_{\mu\nu}=-k(T_{\mu\nu}-\frac{1}{2}g_{\mu\nu}T)$ ，希尔伯特得到的则是 $R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R=-kT_{\mu\nu}$ （两者形式上的差别来自推导方法的截然不同，从而本身也体现了两人在这一环节上的独立性）。需要指出的是，爱因斯坦对场方程的推导在11月25日的论文中写得很简略（ $t_{\mu\nu}$ 在该论文中被称为“能量张量”），在1916年的长篇综述《广义相对论基础》中才详细给出（ $t_{\mu\nu}$ 的名称则改为了“能量分量”）。不过两篇论文的逻辑传承还是很明显的。

(21) 爱因斯坦给洛伦兹的这封信（以及两天后的另一封信）还有两个值得一提的地方：一个是在表示了“推导仍是糟糕的”之后，爱因斯坦开始叙述用作用量原理表述理论的尝试。这一点似乎间接证明了希尔伯特在11月16日的明信片中并未给出场方程。因为否则的话，爱因斯坦写信时由于已得到了正确的场方程，从而也应该已经意识到正确的场方程才与希尔伯特的相一致，以及希尔伯特的作用量是正确的（因为希尔伯特的场方程来自该作用量）。但爱因斯坦给

洛伦兹的这封信却丝毫没有显示出这种知识。另一个是他表示自己没有计算作用量对 $g^{\mu\nu}$ 的变分，因为那太复杂。这说明此类计算在当时绝非是“普通研究生就能完成的常规练习”（参阅第七节）。

(22) 这里所说的“哥廷根数学学会”在文献中有多个英译名，甚至有可能就是第二节中提到过的“哥廷根皇家科学院”。

(23) 关于爱因斯坦写这封信的原因，除回复希尔伯特的友好来信外，有史学家认为还有可能是因为他在写信之前已看到了更强调他的贡献、且援引他所有论文的希尔伯特论文的修改稿，从而情绪有所缓和。另外，这封信的口气似乎意味着两人之间的“不良感觉”已达到了彼此察觉的程度。派斯在《上帝是微妙的》一书中甚至转述了与爱因斯坦有过接触的美国物理学家斯特劳斯（Ernst G. Straus）提供的一条据说得自爱因斯坦的消息，声称那种“不良感觉”曾使希尔伯特给爱因斯坦写过辩解信，表示自己已不记得爱因斯坦的哥廷根演讲。这则消息在我看来有些八卦，因为很难相信希尔伯特会忘记由他亲自请来的爱因斯坦的哥廷根演讲，并以此为由进行辩解。那样的“辩解”不仅太过拙劣，而且只会火上浇油（因为那相当于进一步抹杀了爱因斯坦的贡献），完全不像希尔伯特的风格。

他们为什么反相对论？[\(1\)](#)

一、先声

1919年11月6日，英国皇家学会和皇家天文学会举行了一次联合会议。在会议上，格林尼治天文台（Royal Greenwich Observatory）台长戴森（Frank Dyson, 1868—1939年）报告了英国天文学家爱丁顿（Arthur Eddington, 1882—1944年）对爱因斯坦广义相对论所预言的星光偏折效应的观测检验，结果与广义相对论的预言相一致。爱因斯坦几乎在一夜之间变成了公众和媒体的追逐对象。几星期之后，在应伦敦《泰晤士报》（*Times*）约稿所写的一篇文章中，爱因斯坦对自己的相对论作了一个很幽默的诠释，他说：“如今我在德国被称为德国科学家，在英国则被当作瑞士犹太人。如果哪天我变成一个讨厌鬼的话，那么称呼将会倒过来，在德国我将变成瑞士犹太人，而在英国则将变成德国科学家。”

这位能够预言星光偏折的物理学家或许没有料到，他这段幽默诠释竟然很快就变成了冷酷的事实。那事实就是：20世纪最伟大的物理学家和他最著名的理论遭遇了史上罕见的猛烈攻击——那攻击既发生在欧洲，也发生在东方；既发生在当时，也发生在后世。

让我们从20世纪20年代的柏林（Berlin）开始追溯那一系列发人深思的反相对论运动。[\(2\)](#)

二、德国的反相对论运动

1920年8月24日，距离爱因斯坦那个幽默诠释的发表还不到一年，一场反相对论的公开演讲在柏林最大的音乐厅登场，主讲者是德国实验物理学家格尔克（Ernst Gehrcke，1878—1960年）和铁杆反相对论人士魏兰德⁽³⁾（Paul Weyland，1888—1972年）。一个月后，在德国中部旅游胜地巴登瑙海姆（Bad Nauheim）上演了第二场反相对论公开演讲，主讲者包括了著名的德国实验物理学家、1905年诺贝尔物理学奖得主勒纳德（Philipp Lenard，1862—1947年）。大规模反相对运动由此拉开了序幕。

那些活动开展之时，相对论虽已获得了最具轰动性的观测支持，但毕竟还是一个新理论，而此类活动在当时也还比较新鲜，且主讲者包括了知名的实验物理学家，因此就连爱因斯坦本人也出席了那两次活动，结果却发现那些演讲中没有任何有分量的反对意见，有的只是含糊其辞、捕风捉影的指控和抨击。比如魏兰德和格尔克指控爱因斯坦的某些广义相对论计算剽窃了一位名不见经传的“真正的德国人”的工作，而事实上，那位“真正的德国人”的论文无论前提还是推理都是错误的。勒纳德的反相对论理由则是相对论有违常识，因而必定是错误的。如此浅薄的逻辑出自一位诺贝尔奖得主之口多少有些令人吃惊，以至于爱因斯坦在一份公开声明中表示勒纳德作为实验物理学家虽值得敬重，在理论物理学上却毫无建树，他的反相对论意见更是不值一驳（虽然实际上他还是进行了驳斥）。

不过那些反相对论活动的技术含量虽然很低，它们的公开举行还是令爱因斯坦十分不快，以至于坊间流传起了爱因斯坦打算离开柏林的消息。柏林的对外办公室则收到了来自海外的报告，称那一系列反相对论

事件对德国的国家形象极为不利。受这些消息惊动，主管普鲁士科学、艺术及文化的部长亲自写信给爱因斯坦，谴责了反相对论事件，并表示柏林永远以爱因斯坦为荣。而著名的德国科学家劳厄（Max von Laue, 1879—1960年），能斯特（Walther Nernst, 1864—1941年），鲁本斯（Heinrich Rubens, 1865—1922年）等人则在第一时间公开撰文向爱因斯坦表示了声援。

这些努力虽然暂时挽留住了爱因斯坦，却无法阻止反相对论运动的继续发酵，并逐渐掺入暴力色彩。有人开始宣称，由于运动是相对的，因此打爱因斯坦耳光是合法的，因为那只不过是爱因斯坦的脸颊撞击殴打者的手。更激进的人士甚至主张“让子弹飞”——刺杀爱因斯坦。

在反相对论阵营中，除勒纳德外，还有一位有分量的实验物理学家，那就是1919年诺贝尔物理学奖得主斯塔克（Johannes Stark, 1874—1957年）。⁽⁴⁾反相对论人士除对相对论展开直接攻击外，甚至还对诺贝尔委员会（Nobel Committee）造成了不小的压力，诺贝尔委员会两次请人撰写有关相对论的报告，为给爱因斯坦颁奖做预备，却两次收到了反相对论的报告，左右为难之下只得舍弃相对论，以爱因斯坦在光电效应方面的工作为由给他颁了奖。随着纳粹势力的扩张，德国的反相对论活动也变得更给力，最终于1933年将爱因斯坦赶往了美国。在一些科学史学家眼里，那是一个标志性事件，它标志着世界科学中心由欧洲迁往了美国。

除在德国遭受攻击外，爱因斯坦居住在柏林这一事实，使他很不幸地还同时成为了某些仇视德国的别国人士的攻击对象，甚至在相对中立的美国也不例外。比如某美国刊物的欧洲版曾将爱因斯坦辱骂成猪，因为他是德国人（其实爱因斯坦早就放弃了德国国籍）。而1919年底美国哥伦比亚大学（Columbia University）授予爱因斯坦巴纳德奖章

（Barnard medal）一事，也遭到了一些仇视德国的美国人的抨击。那些事件虽不能与发生在德国的反相对论事件相提并论，却极准确地印证了爱因斯坦的那段幽默诠释。

如今回顾那些陈年旧事时，我们不禁要问：那些人为什么要反相对论？勒纳德和斯塔克早年都曾与爱因斯坦有过很友好的通信，为什么后来却变成了激进的反相对论者？

这其中一个很重要的原因是作为纳粹主义一部分的种族主义。无论勒纳德还是斯塔克，都是忠实的纳粹信徒，主张推进所谓“德意志物理学”（German physics），清除“犹太物理学”（Jewish physics），而声誉鹊起的相对论则被视为是“犹太物理学”的范例——因为爱因斯坦是犹太人。他们的反相对论主张虽然披着物理或哲学的外衣，但在那些不值一驳的表观理由背后所散发的，却往往是种族主义的气息。比如勒纳德反相对论的表观理由虽是相对论有违常识，但很快就将之阴谋化为犹太人蓄意颠覆科学的尝试。而其他一些反相对论者更是公开宣称犹太人的思维已“被永久性地剥夺了对事物本质奥秘的了解”。反相对论者的种族主义攻击不仅针对像爱因斯坦那样的犹太人，而且还波及到了一些“根正苗红”的德国物理学家，比如著名德国物理学家海森伯就因维护相对论而被斯塔克称为了“白色犹太人”（white Jew）。



德国的反爱因斯坦宣传画

除种族原因外，爱因斯坦的政治见解也为他树立了许多敌人。“一战”后的德国因受协约国方面极为严苛的制裁，被迫签订了《凡尔赛条约》（*The Treaty of Versailles*）那样割地加赔款的“不平等条约”，致使国内民族主义（Nationalism）情绪高涨。而爱因斯坦却公开发表了很多和平主义言论，从而成为了民族主义势力的大敌。更糟糕的是，爱因斯坦还因发表过一些同情社会主义的言论，而显著得罪了右翼势力（right wing），被后者斥为是搞“布尔什维克主义者物理学”（Bolshevist physics），以及使物理学沦为辩证法。在主要的反相对论人士中，魏兰德就是右翼人士。此外，在所有知名的反相对论人士中，德国反相对论人士是唯一与爱因斯坦本人存在利益冲突的——因为爱因斯坦居住在德国。有些史学家认为，像勒纳德那样的实验物理学家之所以不遗余力地反相对论，部分原因是爱因斯坦作为理论物理学家的巨大声望威胁到了

实验物理以及他们自身在德国物理学界的地位。

当然，在反相对论人士中，也确实有一些人是出于对相对论这一新理论本身的不适应而反相对论的，比如前面提到的格尔克就是一个例子，他从未加入过纳粹党，而他的反相对论立场则早在纳粹出现前的1911年就已形成，且终生未改其志，直到晚年（1951年）还宣称相对论终有一天将被抛弃。他不仅自己反相对论，还悉心收集了别人的反相对论言论，无形中为史学家们研究那段历史提供了便利。而以相对论有违直觉为由反相对论的人士，也并非都是像勒纳德那样种族主义挂帅的。事实上，当时确有不少人是因担心科学的数学化、抽象化及专门化而反对相对论的，只不过那些人大都是物理学的外行。比如当时有一个自称国家科学院（Academy of Nations）的由哲学家、工程师、医生、退伍军人等外行人士组成的组织，就激烈地反对科学的专门化，试图将科学知识纳入一个统一的、与宗教相容的描述，而相对论则被他们视为头号大敌。

三、苏联的反相对论运动

随着“二战”的结束和纳粹的覆亡，德国的反相对论运动基本消停了，但世界范围内的反相对论运动却还有另外两波。我们前面提到，在德国反相对论运动的核心缘由中，有一条缘由是爱因斯坦同情社会主义，但极具讽刺意味的是，社会主义国家却并不领那个情。事实上，世界范围内的另两波反相对论运动恰恰就都发生在社会主义国家，即苏联^⑤和中国。

这其中发生在苏联的反相对论运动规模较小，系统性的批判主要有两轮：第一轮出现在20世纪30年代的“大清洗”（Great Purge）前后，第二轮则大致始于20世纪40年代后期。1947年，时任苏共领导人之一、主管斯大林时期文化政策的日丹诺夫对相对论宇宙学进行了批评，表示相对论研究者们因不懂辩证法，而错误地把得自于有限经验的规律使用到了无限宇宙中。平心而论，与德国的反相对论运动相比，日丹诺夫的批评虽用语僵化，却还算是比较正经的批评。但在苏联的体制下，日丹诺夫的批评起到了为相关讨论定调的作用，使得更多反相对论的声音随之出炉。不久之后，苏联哲学家卡尔波夫效仿列宁为德国化学家奥斯特瓦尔德（Friedrich W. Ostwald, 1853—1932年）粘贴“伟大的化学家，渺小的哲学家”标签的做法，将爱因斯坦称为是“伟大的科学家和渺小的哲学家”，另一些批判文章则开始采用诸如“反动的爱因斯坦主义”那样更有力度的标签。

与爱因斯坦并无直接瓜葛的苏联为什么要反相对论呢？主要原因在于意识形态，主要战场也在意识形态。爱因斯坦曾多次表示自己在研究相对论的过程中受到过奥地利哲学家马赫（Ernst Mach, 1838—1916年）的影响，然而很不幸的是，在革命导师排行榜上位居探花的“三号

导师”列宁却在《唯物主义与经验批判主义》一书中用很大篇幅批判了马赫，这就使爱因斯坦在苏联的地位变得尴尬。比如马列主义哲学家马克西莫夫就是从这一角度入手，对相对论展开了批判。

不过，相对而言，苏联的反相对论运动无论在规模还是程度上都是比较节制的，而且甚至在斯大林还在世时，就已出现了维护相对论的声音，比如马克西莫夫就被一些苏联物理学家公开斥责为不懂物理。为了维护相对论，苏联学者们还采用了各种迂回的方法。比如个别学者试图论证相对论对欧氏几何的背离是因为受到俄国数学家罗巴切夫斯基（Nikolai Lobachevsky, 1792—1856年）而不是马赫等人的影响。这种论证既弘扬了俄罗斯文化，又替爱因斯坦甩掉了马赫这个包袱，可谓一举两得。可惜这种用心良苦的论证实太过牵强，即便在苏联学者中也鲜有反响。另一类维护相对论的努力则是利用哲学固有的含糊性，试图论证相对论与辩证唯物主义相容。此类努力的代表人物是理论物理学家福克（Vladimir Fock, 1898—1974年），他表示相对论非但不与辩证唯物主义相矛盾，反而是对后者的辉煌证实。而对相对论最有成效的维护，则据说来自苏联核技术的组织者，原子弹、氢弹计划的领导者库尔恰托夫（Igor Kurchatov, 1903—1960年）。据不同渠道的当事人回忆，库尔恰托夫曾向有可能包括斯大林本人在内的苏共高层进谏，表示现代物理是制造原子武器的基础，从而在很大程度上维护了相对论和量子力学的地位——包括其在高校课程中的地位。而原定于1949年召开的用马列主义声讨现代物理的批判大会也因此而取消（一年前，一次针对生物学的类似大会摧毁了苏联的遗传学）。^[6]

1953年之后，随着“四号导师”斯大林的去世，苏联的反相对论运动进一步降温。次年，苏联科学院院士索波列夫（Sergei Sobolev, 1908—1989年）在《真理报》上发表文章，表示对相对论所代表的“唯心主义

世界观”的批判虽是必要的，但试图推翻相对论的物理部分却是徒劳的，从而公开承认了相对论作为物理理论的有效地位。而马克西莫夫的反相对论文章此时则基本没有杂志愿意发表了。

四、中国的反相对论运动

世界范围内的第三波反相对论运动发生在“文革”时期的中国。自1949年之后，中国在意识形态上奉苏联为师，苏联的反相对论思想也经过中国媒体的转载，而传到了中国，并引起了一些山寨版的批判。不过，中国真正大规模、原创性的反相对论运动则是发生在“文革”期间。

1968年，中国科学院成立了一个“批判相对论学习班”^[7]，宗旨是“以毛泽东思想为武器，批判相对论，革相对论的命”，因为“舍此，就不能把自然科学理论推上一个新阶段”。该班成员在其处女作中出手不凡地宣称狭义相对论的光速不变原理“根本违反了唯物辩证法”，意味着“资本主义社会是人类终极社会，垄断资本主义生产力不可超越”。班上的一位年轻学员更是极有创意地指出相对论所鼓吹的“同时的相对性”意味着在不久前发生的中苏冲突中不可能搞明白谁开了第一枪，从而是在袒护苏联。^[8]北京如此，中国的另一重

镇上海也不甘落后，于1969年在复旦大学成立了一个“上海理科批判组”^[9]，宣称“爱因斯坦就是本世纪以来自然科学领域中最大的资产阶级反动学术权威”，“不把相对论之类的反动理论打倒，什么新科学、新技术都是建立不起来的”。



中国的大批判宣传画

与爱因斯坦的瓜葛比苏联更少的中国为什么也要反相对论呢？是学苏联吗？不是。因为“文革”时中苏两国已经交恶，简单地把中国的反相对论运动说成是学苏联无疑是小看了中国人民自己的智慧。事实上，中国的反相对论运动带有鲜明的中国特色——即群众运动色彩。在中国反相对论运动的浪潮中，就连山沟沟里的小学民办教师都有一些被打成了“爱因斯坦黑线”上的人物，可谓是大网垂天、疏而不漏，其波及之广、挖掘之深，皆远非苏联可比，而在言论的荒诞性上，更是连纳粹德国的反相对论运动都要瞠乎其后。

具体地讲，中国的反相对论运动乃是多重原因结出的共同果实。首先当然也是意识形态方面的原因。中苏虽已交恶，但马列著作依然是中国的“圣经”，因此“三号导师”列宁对马赫的批判在中国依然有极大的影响力。其次，中国本土导师毛泽东曾批判过相对主义。这种跟相对论八杆子也打不着的东西落在“批判相对论学习班”的眼里，当场就变成了批判相对论的利器，因为“相对论的大前提是哲学的相对主义”。“上海理科批判组”的悟性也不低，殊途同归地指出了“相对论的要害是相对主义”。这样一来，相对论就与列、毛两大导师的言论同时发生了抵触，在那个一位导师就足以将任何东西送入地狱的年代，它的命运自然可想而知了。

除意识形态方面的原因外，政治斗争也是中国反相对论运动的幕后推手。比如“上海理科批判组”的出现，在很大程度上是当时主持上海工作的张春桥、姚文元两人与北京“中央文革小组”组长陈伯达角力的结果，甚至可以说是三人角力的分战场。张、姚、陈三人对两地的反相对论工作分别作了指导。此外，1972年，借华裔物理学家杨振宁和李政道相继访华之机，当时的中国总理周恩来开始试图恢复中国的科教体系，反相对论运动则因此而成为了极左势力牵制周恩来的舞台之一。

最后，但并非最不重要的，是因为相对论一方面是20世纪最有代表性的物理理论之一，另一方面，却又比与之齐名的量子理论更容易望文生义及粘贴标签，而且相对论创始人爱因斯坦的知名度远大于除他本人外的任何一位量子力学先驱或创始人，因此批判相对论既容易造成影响，又便于新手上路。这一点，以及那个年代无知者无畏、外行批判内行、思辨（如果那也算思辨的话）取代数学、哲学指导科学等风气的盛行，不仅极大地助长了当年的反相对论运动，而且贻害深远，迄今仍是很多民间“科学家”无形的行为准则。

五、结语

最后，让我们小结一下发生在德、苏、中这三个国家的绵延半个多世纪、横跨欧亚两大洲的反相对论运动。从核心缘由上看，反相对论运动兴起的最大因素在德国是作为纳粹主义一部分的种族主义，在苏联是马列主义意识形态，在中国则是马列主义意识形态外加政治斗争。从运动风格上看，德国和苏联的知名反相对论人士大都是不懂相对论的人，他们反相对论的理由虽然荒唐无知，但大都出于自己的真实意愿，是“真心实意”的反对者；而后来居上的中国除了有大批“真心实意”的反对者外，由于动辄搞群众运动式的“人人过关”，还使得一些理解并认同相对论的专业人士被迫发表违心的反相对论言论，成为“口是心非”的反对者。

但总体来说，德、苏、中这三家反相对论的缘由和风格虽各不相同，却有一个共同特点，那就是都倚仗了非科学的手段——无论那手段是纳粹主义、马列主义，还是其他教条——来抨击相对论，从而完全有别于科学探索。我们如今回顾那些反相对论运动的可笑与可悲，并不是要说明相对论不容挑战。诚然，相对论作为一个受到无数观测与实验支持的理论，具有极为坚实的经验基础，但科学上的任何理论都不是金科玉律，相对论也不例外。只不过探索一个科学理论的缺陷，或推翻一个科学理论，必须使用科学的方法，其他手段则大都徒劳无益。那样的手段若是个人使用，导致的通常只是个人闹剧——就像如今仍屡见不鲜的民间“科学家”们的闹剧一样，但若得到国家机器的支持而被强制推行，则会使科学蒙难，国家蒙羞，历史蒙尘。这是我们从那些反相对论运动中最该吸取的教训。

参考文献

[1] Graham L R. Science, Philosophy, and Human Behavior in the Soviet Union [M] . New York: Columbia University Press, 1987.

[2] Grundmann S. The Einstein Dossiers [M] . Berlin: Springer, 2004.

[3] Illy J. Albert Meets America [M] . Baltimore: The John Hopkins University Press, 2006.

[4] Kojevnikov A B. Stalin's Great Science: The Times and Adventures of Soviet Physicists [M] . London: Imperial College Press, 2004.

[5] Levenson T. Einstein in Berlin [M] . New York: Bantam Books, 2004.

[6] Pais A. Subtle is the Lord [M] . Oxford: Oxford University Press, 1982.

[7] Wazeck M. The relativity deniers [J] . New Scientist, 2010, (13) : 48-51.

[8] 胡大年. 爱因斯坦在中国 [M] . 上海: 上海世纪出版集团, 2006.

[9] 周雁翎. 中国自然科学批判运动研究的现状和困境 [J] . 武汉大学学报, 2001.

[10] 柯萨诺夫 V, 维希金 V. 意识形态与核武器——苏联理论物理学的艰难岁月 [J]. 科学文化评论, 2004, 1 (1).

二零一一年一月二十二日写于纽约

(1) 本文的一个只包含德国部分的删节版曾发表于《科学画报》2013年第3期（上海科学技术出版社出版）。

(2) 为行文简洁起见，在下文中我们将用“反相对论”一词来表示反爱因斯坦及反相对论这两种行为。在本文述及的那些事件中这两者几乎总是同时出现的。

(3) 有关魏兰德的生平资料十分有限，此人早年学过化学，后来当过工程师，写过小说，研究过诗歌、雕塑等，皆无建树。他的名声主要来自反相对论。他不仅在德国反相对论，而且还在“二战”后帮美国中情局（CIA）收集反爱因斯坦的材料，算得上是为反相对论事业鞠躬尽瘁了。

(4) 富有戏剧性的是，勒纳德与斯塔克这两位知名反相对论人士都曾做过一些与爱因斯坦存在渊源的工作：勒纳德曾对光电效应（photoelectric effect）做过重要研究，而爱因斯坦获诺贝尔物理学奖的工作正是对光电效应做出理论解释；斯塔克则与爱因斯坦彼此独立地提出了如今被称为斯塔克-爱因斯坦定律（Stark-Einstein law）的光化学定律。

(5) 本文不使用“前苏联”这一称谓，因为并不存在与“前苏联”相对应的“后苏联”，“苏联”一词指的就是“前苏联”，并无歧义。

(6) 史学界对库尔恰托夫在维护相对论与量子力学中所起的作用有一定的争议，因为有关他向苏共高层进谏的说法并无书面记录。

(7) “批判相对论学习班”的全称为“‘批判自然科学理论中资产阶级反动观点’毛泽东思想学习班”，因将批判相对论视为“重中之重”而自我简称为“批判相对论学习班”。

(8) 据许良英先生（《爱因斯坦文集》的主要译者）及其学生屈傲诚的一篇曾摘要发表于《自然辩证法通讯》1984年第6期和1985年第1期上的文章《关于我国“文化大革命”时期批判爱因斯坦和相对论运动的初步考查》披露，此年轻学员乃是郭汉英（郭沫若的儿子，“文革”后曾任中国科学院理论物理研究所研究员、博士生导师等职，于2010年去世）。

(9) “上海理科批判组”的全称为“上海市理科革命大批判写作组”，后期曾以“李柯”为笔名批判相对论。

从爱因斯坦妻子的“秘密贡献”谈起⁽¹⁾

1989年2月，美国物理联合会（American Institute of Physics）主办的《今日物理学》（*Physics Today*）杂志刊登了沃克癌症研究所（Walker Cancer Research Institute）创始人沃克（Evan Harris Walker）的一篇标题为《爱因斯坦用了他妻子的想法吗？》（*Did Einstein Espouse his Spouse's Ideas?*）的文章。在文章中，沃克暗示了一系列颠覆性的观点，其中包括：爱因斯坦的博士论文有可能是妻子米列娃“代笔”的；狭义相对论的基本想法有可能来自米列娃；爱因斯坦有可能销毁了包含这一秘密的信件；爱因斯坦给米列娃的诺贝尔物理学奖的奖金是“封口费”，等等。

那么，爱因斯坦对狭义相对论的贡献是什么呢？沃克认为主要是填补数学细节。稍后，就连这个贡献也遭到了怀疑：1990年，德国女权主义者托梅尔普劳兹（Senta Trömel-Plötz）在一本妇女杂志上发表了一篇题为《米列娃：帮爱因斯坦做数学的女人》（*Mileva Einstein-Marić: The Woman Who Did Einstein's Mathematics*）的文章，认为是米列娃帮助爱因斯坦完成了狭义相对论中的数学。这些文章的作者还暗示或明示，科学史学界对米列娃的贡献进行了有意抹杀。



Mileva Marić（1875—1948年）

别看作者的背景五花八门，结论听起来也荒唐不堪，这类阴谋论观

点却引起了媒体的浓厚兴趣，甚至得到了著名电视公司PBS拍摄的一部有关米列娃的资料片的介绍和传播。就连近几年最流行的爱因斯坦传记《爱因斯坦：他的人生和宇宙》（*Einstein: His Life and Universe*）一书的作者艾萨克森（Walter Isaacson）也称米列娃是“为他（即爱因斯坦）1905年论文提供过数学帮助的塞尔维亚物理学家”。

若上述观点成立，爱因斯坦将变成什么人呢？首先是假博士（因为博士论文是妻子“代笔”的），其次是沽名钓誉之辈（因为狭义相对论的想法和数学都是妻子“赞助”的），再者还是阴险之徒（因为销毁信件，并用金钱封口）。附带着被颠覆的还有整个科学史学界（因为有意抹杀米列娃的贡献）。

有道是：超常的主张需要超常的证据。那么，如此颠覆历史——乃至颠覆史学界——的观点证据何在？主要的证据只有两条：一条是爱因斯坦在1901年给米列娃的信中曾将正在进行的工作称为“我们的”；另一条是已故苏联物理学家约飞（Abram Joffe）1955年曾在某苏联刊物上发文表示见过爱因斯坦的论文署有爱因斯坦和米列娃两人的名字。这其中第一条只是分析性证据，并无一锤定音的效力，而且对信件的排查很快显示，爱因斯坦在热恋时虽偶尔将正在进行的工作称为“我们的”，绝大多数时候仍是用“我的”。不仅如此，米列娃自己在给朋友的信中也称爱因斯坦的工作称为“他的”。可见“我们的”不过是热恋时偶一用之的“甜言蜜语”。更何况，那些信件比爱因斯坦1905年的论文早了足足四年。第二条倒有几分铁证意味，可惜也被查清了，原来约飞只是说论文的作者是一位不知名的专利局职员：爱因斯坦-玛丽克（玛丽克是米列娃的姓），他并且特意说明按瑞士的习惯，妻子的姓氏要添在丈夫姓氏的后面。可见，约飞只是按他认为的瑞士习惯，用“爱因斯坦-玛丽克”来称呼爱因斯坦，而不是说论文上有两个人的名字（事实上他明确说了作者

是一位不知名的专利局职员），更没有说自己曾亲眼见过（事实上有史学家研究了约飞的学术经历后发现，他极不可能见过爱因斯坦论文的预印本）。

如此薄弱的证据居然成就了如此颠覆性的观点，不可谓不荒谬。而如此荒谬的观点，直到2012年仍有史学家撰文驳斥，影响不可谓不持久。其实，这类阴谋论观点在媒体上常能见到，它们的共同特点是：有选择地采用分析性的证据（比如爱因斯坦的一封信件）、不易核实的孤证（比如已故物理学家约飞发表在苏联刊物上的文章），或歪曲臆造的证据（比如对约飞文章的误读），而对不利证据（比如米列娃的数学成绩很差，米列娃的信件未谈及物理，爱因斯坦的工作有自洽而独立的逻辑传承，等等）则要么略而不提，要么进行阴谋论解读（比如毫无证据地认为米列娃的某些信件被爱因斯坦销毁了）。由于现实世界的许多事情都牵涉到许多方面，通过有选择地对待证据，再辅以歪曲臆造等“技术”，往往能轻易构造出颠覆性的观点，而媒体和大众偏偏还特别青睐那样的观点，这是阴谋论观点层出不穷的根本原因。

这原因既成就了阴谋论观点，也提供了识别它的诀窍。虽然普通读者并无能力分辨对证据的歪曲和臆造，但当你看到寥寥几条证据风卷残云般地颠覆一个主流观点，却丝毫不提后者赖以确立的大量其他证据该如何重新解释时，你所看到的就极有可能是一个阴谋论观点。

二零一三年一月二日写于纽约

[\(1\)](#) 本文曾发表于《科学画报》2013年第3期（上海科学技术出版社出版）。

希尔伯特第十问题漫谈⁽¹⁾

一、问题

数学问题是数学中最具魅力的部分之一，并且也是数学史上许多思想和进展的重要源泉。据说有人曾建议德国著名的数学家希尔伯特（David Hilbert, 1862—1943年）去解决费马猜想（Fermat's conjecture），以夺取为这一猜想而设的沃尔夫斯凯尔奖金（Wolfskehl Prize），希尔伯特却笑了笑回答说：“我为什么要杀掉一只会下金蛋的鹅呢？”在希尔伯特看来，一个像费马猜想那样的数学问题对数学的价值是无可估量的。希尔伯特不仅舍不得“杀鹅”，还怀着极大的热诚为20世纪的数学界做了一回“寻鹅之人”。1900年，在巴黎举办的第二届国际数学家大会上，希尔伯特做了一次堪称数学史上影响最为深远的演讲，演讲的题目叫做《数学问题》（*The Problems of Mathematics*）。在那次演讲中，希尔伯特列举了23个他认为最具重要意义的数学问题。⁽²⁾ 那些问题被后人称为“希尔伯特问题”（problems of Hilbert）。自那次演讲之后，解决希尔伯特问题成了许多数学家终生为之奋斗的目标，而在解决这些问题的过程中源源不断产生出来的“金蛋”，则为20世纪的数学发展注入了极大的生机和活力。

在本文中，我们就来讲述有关这些数学问题中第十个——即所谓“希尔伯特第十问题”（Hilbert's tenth problem）——的一些故事。

希尔伯特第十问题是一个与解方程有关的问题。解方程大家都不陌生，在中学时我们就已解过许多简单的方程，比如 $2x-2y=1$ ， $x^2+y^2=z^2$ 等。我们所举的这两个简单方程有一个共同特点，那就是都只包含未知数的整数次幂，而且系数也都是整数，这类方程被称为整系数代数多项

式方程。数学家们对这类方程的研究有着漫长的历史。

在公元3世纪的时候，古希腊数学家丢番图（Diophantus，200?—284?年）发表了一部长篇巨著，叫做《算术》（*Arithmetica*）。这部著作共有13卷，经过1700余年的漫长岁月，目前被公认流传于世的有6卷。⁽³⁾丢番图在这部著作中对整系数代数多项式方程进行了大量研究，那些研究对代数与数论的发展有着先驱性的贡献。后人为了纪念他，把整系数代数多项式方程统称为丢番图方程（Diophantine equation），而丢番图本人，则被一些人尊称为“代数学之父”（the father of algebra）。

对于丢番图方程，数学家们最感兴趣的一个传统问题乃是它是否有整数解（或自然数解）。对于简单的丢番图方程来说，这是很容易找到答案的，比如上面提到的 $x^2+y^2=z^2$ 有整数解（早在3000多年前，中国古代的数学家就知道这个方程的一组特解，即“勾三股四弦五”）；另一方面， $2x-2y=1$ 则没有整数解（因为方程的左边为偶数，右边却为奇数）。但对于一般的丢番图方程来说，判断它是否有整数解却往往是一件极其困难的事情，其中最著名的例子就是上面提到过的费马猜想，即 $x^n+y^n=z^n$ 在 $n>2$ 时没有非零整数解，它是在隔了300多年后才得到的证明。⁽⁴⁾

长期以来，人们对丢番图方程是否有整数解的研究都是针对特定形式的丢番图方程进行的。但是，人们显然也可以提出这样一个问题，即有没有办法对任意形式的丢番图方程是否有整数解进行研究？或者更具体地说，是否能找到一种普遍的算法（algorithm），可用来判定一个任意形式的丢番图方程是否有整数解，从而一劳永逸地解决这类问题？这就是著名的希尔伯特第十问题。这样的问题在数学上被称为判定问题（decision problem），因为它寻求的是对数学命题进行判定的算法。

希尔伯特是一位对数学充满乐观信念的数学家。在他提出希尔伯特第十问题的时候，虽然没有明确表示那样的算法一定存在，但他没有用“是否存在那样的算法”作为问题的表述方式，而是直接要求数学家们寻找那样的算法，可见他对存在一个肯定的答案怀有期待。这种期待与他在其他方面对数学所表示出的乐观看法是一脉相承的。

但是，数学的发展却往往是像希尔伯特那样的数学大师都无法预料的。

二、算法

希尔伯特第十问题要求寻找判定丢番图方程是否有解的算法。但究竟什么是算法呢？在希尔伯特提出问题之时却其实并不存在一个明确定义。这是研究希尔伯特第十问题所遇到的第一个困难。这一困难使得希尔伯特第十问题在提出后整整30年没有取得任何实质进展。

对算法的研究直到20世纪30年代开始才逐渐深入起来。

什么是算法呢？粗略且顾名思义地讲，算法就是（通过有限多的步骤）对数学函数进行有效计算的方法。反过来说，如果一个数学问题能够通过可以有效计算的数学函数得到答案，那么我们就称这一数学问题存在算法。因此算法研究的一个重要的切入点便是寻找可以有效计算的函数。但是，到底什么样的函数是可以有效计算的呢？一开始数学家们并没有普遍结论，只知道一些最简单的函数（比如常数函数），以及用那些函数通过若干简单规则（比如相加）组合出来的函数，是可以有效计算的。数学家们将这类函数称为递归函数（recursive function）。

1931年，年轻的法国数学家赫尔布兰德（Jacques Herbrand, 1908—1931年）对递归函数进行了研究，并给著名逻辑学家哥德尔（Kurt Gödel, 1906—1978年）写信叙述了自己的研究结果。不幸的是，哥德尔当时正处于自己一生最重大的成果——哥德尔不完全性定理（Gödel's incompleteness theorems）——的研究期间，没有立即对赫尔布兰德的工作做出回应。^[5]更不幸的是，那年的夏天，年仅23岁的赫尔布兰德在攀登阿尔卑斯山时不幸遇难，他的工作因此被暂时埋没了起来。

与赫尔布兰德的研究差不多同时，在20世纪30年代初的时候，美国

普林斯顿大学（Princeton University）的逻辑学家丘奇（Alonzo Church, 1903—1995年）也在积极从事逻辑及算法的研究，并且发展出了一套新的逻辑体系。他并且让自己的两位学生——克林（Stephen Kleene, 1909—1994年）与罗瑟（John Rosser, 1907—1989年）——对该逻辑体系做进一步的细致研究。他这两位学生都是第一流的学生，克林更是后来自己也成为了第一流的逻辑学家，他们的研究很快就有了结果，但这结果却大大出乎丘奇的意料：他们发现丘奇的那套体系竟然是自相矛盾的！自相矛盾的逻辑体系只能有一个命运，那就是被放弃。但幸运的是，丘奇的那套体系中有一个组成部分仍然是自洽的，从而可以被保留下来，那个组成部分叫做兰姆达运算（ λ -calculus）。兰姆达运算是做什么用的呢？它可以用来定义函数，而且所有用兰姆达运算所定义的函数都是可以有效计算的。在对兰姆达运算做了研究之后，丘奇提出了一个大胆的猜测，那就是所有可以有效计算的函数都可以用兰姆达运算来定义。

1934年，丘奇向到普林斯顿大学访问的哥德尔介绍了这一猜测，但哥德尔却不以为然。丘奇不服气，于是请哥德尔给出一个他认为合适的方法，来描述可以有效计算的函数。哥德尔没有让他久等，一两个月后就给出了一种完全不同的描述。哥德尔所给出的描述的基础正是3年前赫尔布兰德在给他的信中叙述的结果。哥德尔对这一结果进行了完善。这一结果因此被人们称为赫尔布兰德-哥德尔递归函数。

就这样，丘奇与哥德尔各自给出了一种体系，来描述可以有效计算的函数。那么两者孰是孰非呢？丘奇与克林经过研究，很快证明了这两种看上去完全不同的描述方式实际上是彼此等价的。这两位著名逻辑学家的工作殊途同归大大增强了丘奇的信心，他相信这就算是找到了可以有效计算的函数的普遍定义。但哥德尔及其他一些人对此却仍然怀有疑

虑。

最终打消这种疑虑的是英国数学家图灵（Alan Turing，1912—1954年）。

图灵当时对丘奇及哥德尔在这一领域中的研究并不知情。他所研究的课题是什么样的运算可以用抽象计算机来实现。⁽⁶⁾他的这一研究对后来计算机科学的发展起到了深远的影响。在图灵的研究接近完成的时候，他的导师注意到了丘奇与哥德尔的工作。于是图灵对双方的工作进行了比较，结果发现丘奇与哥德尔所定义的那些函数与他的抽象计算机可以计算的函数恰好吻合！图灵把这一结果作为附录加进了自己的论文中。这一下就连哥德尔也心悦诚服了，毕竟，有什么能比在抽象计算机上可以直接计算更接近“可以有效计算”以及算法的基本含义呢⁽⁷⁾？

在这些有关算法的研究中，数学家们还提出了一个重要的概念，叫做“递归可枚举集”（recursively enumerable set）。什么是递归可枚举集呢？就是那些由可以有效计算的函数所生成的自然数的集合。⁽⁸⁾我们知道，对于集合来说，一个最基本的问题就是判断一个元素是否属于该集合。递归可枚举集也不例外。但是数学家们在研究递归可枚举集的时候，却发现了一个十分微妙的结果，那就是对于某些递归可枚举集来说，我们无法判定一个自然数是否属于该集合！换句话说，有一些递归可枚举集是不可判定的。这一结果把对算法的研究与判定问题联系了起来，为后来解决希尔伯特第十问题埋下了重要的伏笔。

这一系列结果的出现主要集中在1936—1937年间。那时候，在数学中存在无法判定的命题本身已经不是一件新鲜事了。因为早在5年前，哥德尔就已经证明了他那著名的不完全性定理，即任何足够复杂并且自洽的数学体系都必定包含不可判定的命题。⁽⁹⁾但当时已知的不可判定命

题大都集中在逻辑领域内。那么在数学的其他领域中究竟哪些命题是不可判定的呢？这个问题在整整10年之后才开始有答案。

1947年，美国数学家波斯特（Emil Post，1897—1954年）找到了第一个逻辑领域以外的不可判定命题。

波斯特是一位有着深刻洞察力的逻辑学家，7岁时随父母从波兰移民到美国，是美国逻辑学领域的先驱者之一。他比哥德尔早了将近10年就得到了与哥德尔不完全性定理相似的结果，但由于想对结果作更全面的分析而没有及时发表。1936年，几乎与上面提到的哥德尔、丘奇及图灵同时，波斯特独立提出了非常类似于图灵的结果。波斯特同时还是最早意识到希尔伯特第十问题可能会有否定答案的数学家之一。他在1944的一篇文章中明确提出希尔伯特第十问题“期待一个不可解性证明”。当时波斯特在纽约市立大学（The City College of New York）任教，他的这一观点深深地打动了一位学生，使后者走上了毕生钻研希尔伯特第十问题的征途。

那位学生名叫戴维斯（Martin Davis，1928— ），是解决希尔伯特第十问题的关键人物。

三、丢番图集

戴维斯的父母也是从波兰移民来美国的，戴维斯本人则出生在纽约。1944—1948年间，戴维斯在纽约市立大学学习，波斯特对希尔伯特第十问题“期待一个不可解性证明”的看法用戴维斯本人的话说是开始了他“对这一问题的终身迷恋”。从纽约市立大学毕业后，戴维斯来到了美国逻辑学的中心普林斯顿，跟随丘奇从事进一步的研究。戴维斯在普林斯顿研究的是一个冷门课题——别小看这种课题，对研究生来说，研究那样的课题往往最容易出成果，且不容易被别人抢先或与人“撞车”。但戴维斯无法抵御希尔伯特第十问题的魅力，在研究自己课题的同时，还分出精力来继续思考希尔伯特第十问题。最后他甚至在博士论文上特意增添了一个章节，简单叙述了自己在希尔伯特第十问题上“不务正业”的结果，那是在1950年。这一增添的章节使戴维斯的那篇原本会像多数研究生工作那样被人遗忘的博士论文变得重要起来，甚至可以名垂史册。3年后，戴维斯发表了一篇更详细的论述。他的这一工作标志着数学家们正式开始解决希尔伯特第十问题。

戴维斯在他的研究中引进及应用了一个很重要的概念，叫做“丢番图集”（Diophantine set）。和我们在前面提到过的递归可枚举集一样，丢番图集也是一些由自然数组成的集合。所不同的是，递归可枚举集是由可以有效计算的函数生成的，而丢番图集则是通过丢番图方程生成的。戴维斯的重要发现就在于找到了这两类集合之间的一种关联。

戴维斯为什么要引进丢番图集的概念呢？是因为倘若希尔伯特第十问题具有肯定的答案，即存在一个算法来判定丢番图方程是否有解，那么我们就可以用这一算法来确定一个自然数是否属于某个丢番图集，这表明所有丢番图集都是可判定的。反过来说，倘若我们可以证明某些丢

番图集是不可判定的，也就等于证明了希尔伯特第十问题具有否定的答案。因此，丢番图集的可判定与否跟希尔伯特第十问题具有肯定还是否定答案有着密切关系，这正是戴维斯引进丢番图集这一概念的价值所在。

受波斯特影响，戴维斯倾向于认为希尔伯特第十问题具有否定答案，因此，他想要证明的是某些丢番图集是不可判定的。那么，怎样才能证明某些丢番图集是不可判定的呢？最好的办法就是设法把它与某一类已经知道是不可判定的集合联系在一起。⁽¹⁰⁾而什么样的集合是已经知道是不可判定的呢？正是前面提到过的递归可枚举集。因此在这两类集合之间建立关联是非常重要的。尤其是，如果能证明所有递归可枚举集都是丢番图集，那就等于证明了某些丢番图集是不可判定的，从而也就等于证明了希尔伯特第十问题的否定答案——即波斯特所说的“不可解性证明”。

这正是戴维斯寻找希尔伯特第十问题否定答案的思路。

不幸的是，在戴维斯找到的丢番图集与递归可枚举集的关联中用到了一个被称为“有界全称量词”（bounded universal quantifier）的逻辑算符。如果没有这个有界全称量词，他就可以证明所有递归可枚举集都是丢番图集，一切也就大功告成了。可是数学证明是差不得分毫的，因为有了这个有界全称量词，戴维斯的逻辑链条中断了，从而无法对希尔伯特第十问题作出解答。但尽管如此，戴维斯仍然相信所有递归可枚举集都是丢番图集，他把这一点作为一个猜测提了出来。

在当时，这是一个很大胆的猜测。

很明显，要证明戴维斯的猜测，关键就得把那个有界全称量词去

掉，这是一件非常困难的事情。直到9年之后，即1959年，戴维斯才在与美国哲学家普特南（Hilary Putnam, 1926— ）的合作中有条件地做到了这一点。他们为做到这一点所付出的代价，是不得不引进了两条额外假设。

初看起来，这像是不进反退：原本只有一个麻烦（即“有界全称量词”），现在反而变成了两个（即“两条额外假设”）。但数学假设的证明难度不是用数量来简单衡量的，戴维斯与普特南所引进的那两条额外假设比那个有界全称量词来得具体，因而处理起来要容易一些。在发表这一研究的全文之前，戴维斯与普特南把自己的结果寄给了研究希尔伯特第十问题的另一位重要人物——美国数学家罗宾逊夫人（Julia Robinson, 1919—1985年），想先听听她的看法。

这一寄揭开了一段新的合作，把他们的结果又大大向前推进了一步。

四、罗宾逊猜想

数学是一个性别比例相当失调的领域，罗宾逊夫人是该领域中为数不多的女数学家之一。与其他一些女数学家一样，她一生在追求学术的过程中遇到过许多坎坷。这些坎坷既有来自社会的，也有来自自己生活的。罗宾逊夫人幼年时屡患疾病，导致身体虚弱，无法生育，这一点曾使酷爱家庭的她陷入极度的痛苦之中。后来，在她那同为数学家的丈夫引导下，是数学的力量让她渐渐摆脱了痛苦的阴影。罗宾逊夫人的丈夫早年曾是她的数论教授，帮助她打下了非常扎实的数论基础。

罗宾逊夫人自1948年起开始研究希尔伯特第十问题，并曾与戴维斯有过交流。当她收到戴维斯与普特南寄来的结果时，凭借自己的数论功底，很快就发现了他们所作的两个假设中有一个可以去掉，同时整个证明也可以作极大的简化。1961年，戴维斯、普特南及罗宾逊夫人合作发表了这一简化后的结果。这一结果是戴维斯、普特南的逻辑技巧与罗宾逊夫人的数论功底的完美结合，也是数学家们在希尔伯特第十问题的研究中所取得的又一个重要进展。

在这一进展中戴维斯与普特南所作的两个假设只剩下了一个。但剩下的一个却是连罗宾逊夫人也无法去除的。这个假设涉及一种被称为“指数丢番图集”（exponential Diophantine set）的集合，这种集合类似于丢番图集，但却涉及指数函数。倘若有人能证明指数丢番图集实际上就是丢番图集——这也正是那剩下的假设之所在，那么戴维斯、普特南及罗宾逊夫人的工作就完全了，希尔伯特第十问题也就被证明具有否定答案了。但指数丢番图集究竟是不是丢番图集呢？这个问题困住了这三个人。

对罗宾逊夫人来说，指数丢番图集其实并不陌生。早在1948年，当她刚刚涉足希尔伯特第十问题的时候，就研究过由著名波兰逻辑学家塔尔斯基（Alfred Tarski, 1901—1983年）所提出的一个猜测。该猜测认为指数丢番图集不是丢番图集。经过一段时间的研究后，罗宾逊夫人开始怀疑起了塔尔斯基的猜测，因为她找不到任何证据可以支持这一猜测。于是她转而作出了一个与塔尔斯基相反的猜测，即指数丢番图集实际上就是丢番图集，这个命题被称为罗宾逊猜想。这也正是戴维斯、普特南及罗宾逊夫人1961年的工作中唯一缺失的环节。他们距离希尔伯特第十问题的解决只剩下了一步之遥，但这一步却难似登天。

在罗宾逊夫人沉醉于希尔伯特第十问题的那些年里，幼年患病留下的后遗症一再困扰着她。当年的一位医生甚至预言她的心脏机能受损严重，也许活不过40岁。这一预测虽然很幸运地由于后来的一次成功的心脏手术而没有成为事实，但每一年的生日，罗宾逊夫人都要在吹熄蜡烛的时候许愿，希望能够看到希尔伯特第十问题的解决——无论谁来解决都可以，但一定要在她有生之年解决。“我无法忍受在不知道答案的情况下离开人世”——这是罗宾逊夫人的话。

时光一年一年地流逝，罗宾逊夫人的愿望一次又一次地落空。那手握最后一把钥匙的人究竟在哪里呢？

在那些年，戴维斯也常常被人问到这一问题。当时正是“冷战”时期，对美国人来说世界上最遥远的地方莫过于是俄国。因此，戴维斯总是戏剧性地回答说：“那会是一位聪明的俄国年轻人。”如果戴维斯是一位占星师的话，这句回答足可让他震动天下，因为他每一个字都说对了！一位聪明的俄国年轻人从世界的另一端走上了数学舞台，他的名字叫做马蒂亚塞维奇（Yuri Matiyasevich, 1947— ），他将为这根长长的智慧链条扣上最后一环。

五、解决

马蒂亚塞维奇1947年出生在苏联的列宁格勒（俄罗斯的圣彼得堡）。他12岁时父亲就不幸去世，但家境贫寒的马蒂亚塞维奇凭借优异的数学成绩在苏联的数学竞赛体系中脱颖而出，获得了良好的教育机会。1965年，在他还在念本科的时候，他的导师马斯洛夫（S. Yu. Maslov, 1939—1982年）就建议他去证明丢番图方程的不可判定性。马斯洛夫在建议时轻描淡写地补充说“这个问题也被称为希尔伯特第十问题，但你不必理会这个”。马蒂亚塞维奇说他对研究这类不可解问题没有经验，马斯洛夫回答说不可解问题没什么大不了的，无非就是把它约化成一个已知是不可解的其他问题。他还告诉马蒂亚塞维奇说有几个美国人曾做过一些研究，但不必理会那些研究，因为它们“很可能是不充分的”。



带着马斯洛夫的建议，马蒂亚塞维奇开始研究起了希尔伯特第十问题。但他的研究并不顺利，他一度误以为已解决了问题，甚至开始准备做报告，结果却发现自己犯了一个错误。经过了一段时间的徒劳无功之

后，他开始阅读起“几个美国人”的那些“很可能是充分的”工作来，但依然没有获得实质性进展，倒是他作为“研究希尔伯特问题的本科生”的名声走红了校园，不时遭来一些善意的取笑。毕业的时间渐渐临近，他只好把这个艰深的问题放在一边，以便可以有时间做一些其他的工作——比如应付毕业论文。

一晃又是几年，到了1969年，顽强的罗宾逊夫人又向希尔伯特第十问题发起了一次冲击。这一次虽然还是没有成功，但她为证明罗宾逊猜想提出了一条非常巧妙的思路。罗宾逊夫人的结果发表后，很快有同事把这一消息告诉了马蒂亚塞维奇。但这时的马蒂亚塞维奇早已决定不再把时间浪费在希尔伯特第十问题上了，于是没有理会这一消息。不过，事情接下来的发展变得很富有戏剧性，用马蒂亚塞维奇自己的话说：“在数学天堂的某个角落里必定存在着一位数学之神（或女神），不想让我错过罗宾逊夫人的新论文。”由于他此前对希尔伯特第十问题的研究，苏联的一份数学评论杂志把罗宾逊夫人的论文寄给了他，让他加以评论。就这样，马蒂亚塞维奇终于还是看到了罗宾逊夫人的论文。

这一看之下他立刻被罗宾逊夫人的思路所吸引，重新投入到了希尔伯特第十问题的研究上来。

在接下来的几个月时间里，马蒂亚塞维奇一直在思索罗宾逊猜想。1969年在不知不觉间落下了帷幕。在除夕夜的派对上，马蒂亚塞维奇因过于出神，走的时候竟错穿了他叔叔的衣服离去。这样全神贯注的投入终于获得了巨大的成功。1970年新年到来后的第4天，马蒂亚塞维奇成功地证明了罗宾逊猜想，从而一举解决了希尔伯特第十问题。但有了几年前误以为解决希尔伯特第十问题的教训，这一次他把文章交给了马斯洛夫及另一位数学家栗弗席茨（Vladimir Lifshits），请他们检验，然后携未婚妻出外滑雪度假。两个星期后当他回到学校时，一切都变了，他

的论文经受住了以眼光犀利著称的数学家法蒂夫（D. K. Faddeev, 1907—1989年）与马尔科夫（A. A. Markov, 1903—1979年）的检验。

马蒂亚塞维奇成为了希尔伯特第十问题的解决者。

1月29日，马蒂亚塞维奇做了有关他研究成果的第一次公开演讲。那次演讲中的一位听众把这一成果带到了不久之后在新西伯利亚（Novosibirsk）召开的一次数学会议上，而那次会议的出席者中恰好有一位是罗宾逊夫人的同事。就这样，马蒂亚塞维奇解决希尔伯特第十问题的消息很快传遍了数学界。那时候马蒂亚塞维奇还不满23岁，正是一位“聪明的俄国年轻人”。

2月25日，罗宾逊夫人接到了同事的电话，告知她这一消息。那一年的生日，当罗宾逊夫人又将吹熄生日蜡烛时，她停了下来，忽然意识到自己许了这么多年的愿望已经成为了现实，那是一种美妙的感觉。虽然她自己曾经那么地接近答案，却还是失之交臂，但她没有觉得遗憾，因为对于像她那样真正热爱数学的人来说，对数学真理的欣赏远远超越了任何个人的荣誉。在给马蒂亚塞维奇的祝贺信中，罗宾逊夫人这样写道：“让我特别高兴的是，当我想到我最初提出那个猜想的时候，你还是个孩子，而我不得不等待你的长大。”戴维斯也非常兴奋，他在自己的经典著作《可计算性与不可解性》（*Computability and Unsolvability*）的平装本序言中写道：“我一生最大的快乐之一是1970年2月读到马蒂亚塞维奇的工作”。而年轻的马蒂亚塞维奇同样对戴维斯、罗宾逊夫人，以及在解决希尔伯特第十问题的漫长征途中做出贡献的所有前辈数学家表达了深深的敬意。

在20世纪六七十年代那个寒冷的政治冬天里，这些第一流的数学家

他们用他们的杰出工作划开了“冷战”的冰层，让世人看到了科学的伟大人文力量。按照罗宾逊夫人的说法，这是一种存在于科学家心中的观念，它跨越地理、种族、意识形态、性别、年龄甚至时代而存在，过去、现在及未来的所有数学家们彼此都是同事，他们献身于一个共同的目标，那便是最美丽的科学与艺术。

这是希尔伯特第十问题留给我们最丰厚的精神遗产。

二零零五年八月二十四日写于纽约

附录：丢番图方程

本附录将对希尔伯特第十问题中的一个基本概念——丢番图方程——做简单的介绍，并证明对于研究希尔伯特第十问题来说，把丢番图方程的解限定在整数、正整数及自然数（零及正整数）上是彼此等价的。[\(11\)](#)这一点我们在第一节的注释（即第134页注①）中提到过，但未予证明。

我们从希尔伯特第十问题的原始表述——即希尔伯特本人的表述——开始讲起：

希尔伯特第十问题：给定一个具有任意多个未知数的整系数丢番图方程：寻找一个可以通过有限多次运算（operation）确定该方程是否有整数解的程序（process）。

这里出现的第一个数学概念就是丢番图方程。如我们在第一节中介绍的，这一名称是为了纪念古希腊数学家丢番图[\(12\)](#)。丢番图方程是指未知数只取整数值的代数多项式方程，它的一般形式为

$$P(x_1, x_2, \dots, x_k) \equiv \sum a_{i_1 \dots i_k} x_1^{i_1} \cdots x_k^{i_k} = 0$$

式中， x_1, x_2, \dots, x_k 为未知数， $a_{i_1 \dots i_k}$ 为常系数，求和对 i_1, i_2, \dots, i_k 进行。在希尔伯特第十问题中，丢番图方程的所有系数 $a_{i_1 \dots i_k}$ 都是整数，这样的方程叫做整系数丢番图方程。[\(13\)](#)

毫无疑问，具体的丢番图方程的数目是无穷的。丢番图方程的繁简及难易程度可以天差地别：简单丢番图方程的求解或判断其是否有解是连中学生都可以做到的；而复杂的丢番图方程则可以复杂到令人难以想象的地步，复杂到就算把全世界所有的纸都用上，也无法写下的程度——因为丢番图方程中未知数的个数、它们的最高幂次都是任意的。在很多时候，一个丢番图方程甚至不需要有很多的未知数，不需要有很大的系数，也不需要有很高的幂次，就能使得判定其是否有解变得极其困难。⁽¹⁴⁾丢番图方程的复杂和艰深，使之成为了数论中一个专门的研究领域，被称为丢番图分析（**Diophantine analysis**）。

有些读者可能会问：既然对特定丢番图方程（比如费马大定理所涉及的丢番图方程）是否有解的判定就已如此困难，动辄花费数学家们几百年的时间，那希尔伯特要大家研究一般丢番图方程是否有解，其胃口是不是太大了一点？的确，希尔伯特把这个问题列入他的讲稿曾经令一些数学家感到意外。但是，事情——尤其是数学上的事情——有时就是那么奇妙，把一大堆极其困难的个体问题合在一起，研究其群体的性质有时反而会有意想不到的优势。当然，希尔伯特是否是因为考虑到了这一点而“狮子大开口”，后人不得而知，但希尔伯特第十问题的最终解决似乎印证了这种有意思的情形。⁽¹⁵⁾

在希尔伯特的表述中，他要求大家寻求的是确定丢番图方程是否有整数解的程序。但后世的数学家们在研究希尔伯特第十问题的时候，通常把解限制在正整数或自然数的范围内。为什么可以作这种限制呢？我们现在就来证明一下。

我们首先可以证明，倘若存在一个确定丢番图方程是否有自然数解的程序，则必定也存在一个确定丢番图方程是否有整数解的程序。这是

因为要确定丢番图方程 $P(x_1, x_2, \dots, x_k) = 0$ 是否有整数解，只要逐一检验 2^k 个丢番图方程（每个未知数带正负两种可能的符号， k 个未知数共计有 2^k 种组合）：

$$P(x_1, x_2, \dots, x_k) = 0$$

$$P(-x_1, x_2, \dots, x_k) = 0$$

.....

$$P(x_1, x_2, \dots, -x_k) = 0$$

是否有自然数解即可，如果这 2^k 个丢番图方程中任何一个有自然数解，则 $P(x_1, x_2, \dots, x_k) = 0$ 必定有整数解；反之，倘若这 2^k 个丢番图方程没有一个有自然数解，则 $P(x_1, x_2, \dots, x_k) = 0$ 必定没有整数解。因此，存在一个确定丢番图方程是否有自然数解的程序，就必定也存在一个确定丢番图方程是否有整数解的程序。

然后我们再证明，倘若存在一个确定丢番图方程是否有整数解的程序，则必定也存在一个确定丢番图方程是否有自然数解的程序。这里我们要用到一个简单的数学定理，即拉格朗日四平方定理（Lagrange's four-square theorem）。这个定理表明任何一个自然数都可以表示成不超过四个整数的平方之和。运用这个定理，我们可以把对 $P(x_1, x_2, \dots, x_k) = 0$ 是否存在自然数解的判定约化为对包含 $4k$ 个自变量 a_1, \dots, d_k 的丢番图方程

$$P(a_1^2 + b_1^2 + c_1^2 + d_1^2, a_2^2 + b_2^2 + c_2^2 + d_2^2, \dots, a_k^2 + b_k^2 + c_k^2 + d_k^2) = 0$$

是否存在整数解的判定。因此存在一个确定丢番图方程是否有整数解的程序，就必定也存在一个确定丢番图方程是否有自然数解的程序。

这样，我们就证明了对于研究希尔伯特第十问题来说，把丢番图方程的解限定在整数及自然数上是彼此等价的。将这一等价性扩展到包含正整数是极其容易的（请读者自行完成）。这样我们就证明了在研究希尔伯特第十问题（或介绍希尔伯特第十问题）时把解限制在自然数或正整数范围内的合理性。

在本附录的最后，我们还要来证明这样一点：那就是如果存在一个确定丢番图方程在某个范围（比如整数、正整数及自然数等）内是否有解的程序，则同一程序也可以用来确定任意丢番图方程组在同一范围内是否有解。证明很简单：因为丢番图方程组

$P_i(x_1, x_2, \dots, x_k) = 0 (i = 1, 2, \dots, n)$ 在某一范围内有解的充要条件为丢番图方程 $P_1^2 + P_2^2 + \dots + P_n^2 = 0$ 在同一范围内有解。因此，人们在对丢番图方程做一般讨论的时候，事实上已经把丢番图方程组也包含在内了。

二零零五年十月十三日写于纽约

(1) 本文的正文部分曾发表于2005年10月26日和11月2日《中国青年报》的“冰点探索”栏目，发表时的标题分别为《会下金蛋的鹅——希尔伯特第十问题（上）》和《会下金蛋的鹅——希尔伯特第十问题（下）》。

(2) 这23个问题中有一些——比如第八问题——不是单一问题。另外，这些问题是以演讲的文稿而非演讲本身为依据排列的，希尔伯特在演讲中直接提及的只有其中的10个问题（本文所述的第十问题不在其中）。

(3) 除公认的6卷外，另有4卷发现于20世纪的阿拉伯文抄本也被认为有可能是丢番图《算术》或其注释本的译本。

(4) 细心的读者可能会注意到，在本文中我们没有对整数、正整数及自然数（零及正整数）等做出区分。这是因为可以证明（参阅附录），对于希尔伯特第十问题来说，把解限定在这些数集的任意一者中都是等价的。

(5) 哥德尔给赫尔布兰德的回应只是不够“立即”，而非没有。他的回信写于1931年7月25日，赫尔布兰德遇难的时间则是7月27日，只相隔了两天，赫尔布兰德没来得及收到回信就去世了。

(6) 具体地讲，图灵当时的目的是要研究希尔伯特于1928年提出的有关一阶逻辑的判定问题。

(7) 不过，需要提醒读者的是，把可以有效计算的函数等同于丘奇、哥德尔、图灵等人提出的、彼此等价的函数并不是建立在数学证明的基础之上的，而只是一个猜测性的论题，即所谓“丘奇论题”（Church's thesis）。一般认为，丘奇论题是无法被证明的，因为“有效计算”本身是一个不存在精确定义的概念，它本质上取决于人们对“有效”及“计算”这样的非精确概念的理解。如果有一天人们发现有必要改变或拓展原先对这些概念的理解，则数学上的一些相关结果——包括希尔伯特第十问题的解决方式——有可能会需要作出相应的改变。

(8) 读者们想必猜到了，这些集合之所以被称为“递归可枚举集”，乃是因为可以有效计算的函数的定义之一叫做“赫尔布兰德-哥德尔递归函数”。

(9) 确切地讲，这是哥德尔第一不完全性定理（Gödel's first incompleteness theorem）。

(10) 这里的用词需稍作说明：当我们对集合使用“不可判定”一词时，有两种不同的用法（读者不难依上下文做出区分）：第一种是针对某个特定的集合，指的是无法判定一个自然数是否属于该集合；第二种是针对某一类集合，指的是该类集合中至少有某些特定集合是在第一种用法下不可判定的。

(11) 数学界对自然数是否包含零没有统一的约定，在本文中我们用它来表示零和正整数，不习惯这一用法的读者可以自行改用“非负整数”这一术语。

(12) 丢番图通常被称为古希腊数学家，但他究竟是不是古希腊人其实略有争议——有人认为他也许是巴比伦人（Babylonian）、犹太人（Jewish）或迦尔底亚人（Chaldean）。

(13) 人们提到丢番图方程时通常指的就是整系数丢番图方程，因此“整系数”这一限定语常被省略。

(14) 比如费马大定理只含有三个未知数，系数皆小，却花费了数学家们300多年的时间才得以证明；欧拉四次方假设（Euler quartic conjecture）——它猜测 $x^4+y^4+z^4=w^4$ 没有正整数解——所含的幂次只有四次，系数也皆小，却花费了数学家们200多年的时间才得以否认。

(15) 不过，希尔伯特第十问题的解决并不意味着解决了所有具体丢番图方程的求解或判定其是否有解的问题，这一点是需要注意的。

黎曼猜想浅说(1)

一

2000年5月24日，美国克雷数学研究所（Clay Mathematics Institute）在法国巴黎召开了一次数学会议。在会议上，与会者们列出了七个数学难题，并作出了一个颇具轰动性的决定：为每个难题设立100万美元的巨额奖金。距此次会议100年前的1900年，也是在巴黎，也是在一次数学会议上，一位名叫希尔伯特（David Hilbert）的德国数学大师也列出了一系列数学难题。那些难题一分钱的奖金都没有，但对后世的数学发展产生了深远影响。这两次远隔一个世纪遥相呼应的数学会议除了都在巴黎召开外，还有一个共同之处，那就是在所列出的难题之中，有一个——并且只有一个——是共同的。

那个难题就是“黎曼猜想”（Riemann hypothesis）。



黎曼猜想顾名思义，是由一位名叫黎曼（Bernhard Riemann）的数学家提出的，那位数学家于1826年出生在如今属于德国，当时属于汉诺威王国（Kingdom of Hanover）的一座名叫布列斯伦茨（Breselenz）的小镇。1859年，黎曼被选为了柏林科学院的通讯院士。作为对这一崇高荣誉的回报，他向柏林科学院提交了一篇题为《论小于给定数值的素数

个数》的论文。那篇只有短短八页的论文就是黎曼猜想的“诞生地”。

黎曼那篇论文所研究的是一个数学家们长期以来就很感兴趣的问题，那就是素数的分布。素数是像2、5、19、137那样除了1和自身以外不能被其他正整数整除的数。这些数在数论研究中有着极大的重要性，因为所有大于1的正整数都可以表示成它们的乘积。从某种意义上讲，它们在数论中的地位类似于构筑万物的原子在物理世界中的地位。素数的定义简单得可以在中学，甚至小学课上进行讲授，但它们的分布却奥妙得异乎寻常，数学家们付出了极大的心力，却迄今未能彻底了解。黎曼那篇论文的一个重大成果，就是发现素数分布的奥秘完全蕴藏在一个特殊的函数之中——尤其是，使那个函数取值为零的一系列特殊的点对素数分布的细致规律有着决定性的影响。那个函数如今被称为黎曼 ζ 函数，那一系列特殊的点则被称为黎曼 ζ 函数的非平凡零点（下文中有时将简称其为零点）。

有意思的是，黎曼那篇论文的成果虽然重大，文字却极为简练，甚至简练得有些过分，因为它包括了很多“证明从略”的地方。而要命的是，“证明从略”原本是该用来省略那些显而易见的证明的，黎曼的论文却并非如此，他那些“证明从略”的地方有些花费了后世数学家们几十年的努力才得以补全，有些甚至直到今天仍是空白。

黎曼为什么要把那么多并非显而易见的证明从略呢？我们无法确知，也许是因为它们对于他来说确实是显而易见的，也许是因为不想花太多时间来撰写文章。但有一点基本可以确定，那就是他的“证明从略”绝不是类似于调皮学生蒙混考试的做法，而且很可能也并不是把错误证明当成正确的盲目乐观——后者在数学史上不乏先例，比如法国数学家费马（Pierre de Fermat）在写下费马猜想时所表示的“我发现了一个真正出色的证明，可惜页边太窄写不下来”就基本已被数学界认定是把

错误证明当成正确的盲目乐观。因为人们后来从黎曼的手稿中发现他对许多论文中从略了的证明是做过扎实研究的，而且那些研究的水平之高，甚至在隔了几十年之后被整理出来时，有时也仍具有极大的领先性。

但黎曼的论文在为数不少的“证明从略”之外，却引人注目地包含了一个他明确承认自己无法证明的命题，那个命题就是黎曼猜想。

那么，黎曼猜想究竟是一个什么猜想呢？简单地说，是一个关于我们前面提到的，对素数分布的细致规律有着决定性影响的黎曼 ζ 函数的非平凡零点的猜想。关于那些非平凡零点，容易证明的结果只有一个，那就是它们都分布在一个带状区域上，但黎曼认为它们的分布要比这个容易证明的结果齐整得多，他猜测它们全都位于该带状区域正中央的一条直线上，这就是所谓的黎曼猜想。而这条被猜测为包含黎曼 ζ 函数所有非平凡零点的直线则被称为临界线。

二

黎曼猜想自1859年“诞生”以来，已经过了一百五十多个春秋。在这期间，它就像一座巍峨的山峰，吸引了无数数学家前去攀登，却谁也没能登顶。当然，如果仅从时间上比较的话，黎曼猜想的这个纪录跟费马猜想时隔三个半世纪以上才被解决，以及哥德巴赫猜想历经两个半世纪以上仍屹立不倒相比，还差得很远。但黎曼猜想在数学上的重要性却要远远超过这两个大众知名度更高的猜想。有人统计过，在当今数学文献中已有超过一千条数学命题以黎曼猜想（或其推广形式）的成立为前提。这意味着：如果黎曼猜想及其推广形式被证明，所有那些数学命题就全都可以荣升为定理；反之，如果黎曼猜想被否定，则那些数学命题中起码有一部分恐将成为陪葬。一个数学猜想与为数如此众多的数学命题的命运息息相关，是极为罕有的。

不过，数学家们攀登黎曼猜想这座巍峨山峰的努力虽迄今未能取得完全成功，在这过程中却也取得了一些阶段性成果，好比是扎下了几座营寨。

这其中第一个阶段性成果出现在黎曼猜想问世37年后的1896年。我们在前面提到过，关于黎曼 ζ 函数的非平凡零点，容易证明的结果只有一个，那就是它们都分布在一个带状区域上。那个阶段性成果是什么呢？就是将那个带状区域的边界剔除了——也就是说，黎曼 ζ 函数的非平凡零点只分布在那个带状区域的内部，而不包括边界。这个成果是由法国数学家哈达玛（Jacques Hadamard）与比利时数学家普森（Charles de la Vallée-Poussin）彼此独立地取得的。

粗看起来，这似乎是很微不足道的成果，一个带状区域的边界跟它

的内部相比，从面积上讲比例实际上是零。但是别小看了这个成果，它对于研究黎曼猜想来说只是一小步，对于研究另一个数学猜想来说却是巨大的飞跃，因为它直接导致了后者的证明。那个数学猜想如今已被称为素数定理（prime number theorem），它所描述的是素数的大范围分布规律。素数定理自被提出以来悬而未决已超过一百年，在当时乃是一个比黎曼猜想更令数学界期待的东西。

在上述成果之后又隔了18年，1914年，丹麦数学家玻尔（Harald Bohr）与德国数学家兰道（Edmund Landau）取得了另一个阶段性成果，那就是证明了黎曼 ζ 函数的非平凡零点倾向于“紧密团结”在临界线的周围。这个结果用数学语言来说，就是包含临界线的无论多么窄的带状区域都包含了黎曼 ζ 函数的几乎所有的非平凡零点。不过“紧密团结”归“紧密团结”，这一结果却不足以证明任何一个零点恰好就在临界线上，因此它距离黎曼猜想的要求仍然相差很远。

但就在那同一年，另一个阶段性成果出现了：英国数学家哈代（Godfrey Hardy）终于将“红旗”插上了临界线——他证明了黎曼 ζ 函数有无穷多个非平凡零点位于临界线上。粗看起来，这似乎是一个非同小可的结果，因为黎曼 ζ 函数的非平凡零点总共就是无穷多个，而哈代证明了有无穷多个零点位于临界线上，从字面上看，两者已经一模一样了。可惜的是，“无穷”乃是数学中一个很微妙的概念，同样是无穷，彼此却未必是一回事，不仅未必是一回事，简直可以要差多远就差多远，甚至差无穷远！因此，为了知道哈代的结果离黎曼猜想的要求还有多远，我们需要更具体的结果。

那样的具体结果出现在7年后的1921年。那一年，哈代与英国数学家李特伍德（John Littlewood）合作，对自己7年前那个结果中的“无穷”做出了具体估计。那么，按照他们的具体估计，那已被证明为位于

临界线上的“无穷多个非平凡零点”跟全部非平凡零点相比，究竟占多大的百分比呢？答案可能沮丧得出乎读者们的意料：百分之零！

数学家们将这个百分比推进到一个大于零的数字是在21年后的1942年。那一年，挪威数学家赛尔伯格（Atle Selberg）终于证明了这个百分比大于零。赛尔伯格做出这项成果时正值第二次世界大战的硝烟在欧洲各地弥漫，他所在的挪威奥斯陆大学几乎成了一座孤岛，连数学期刊都无法送达。但赛尔伯格并不在乎，他表示“这就像处在一座监狱里，你与世隔绝了，但你显然有机会把注意力集中在自己的想法上，而不会因其他人的所作所为而分心，从这个意义上讲我觉得那种情形对于我的研究来说有许多有利的方面”。赛尔伯格很好地利用了那“许多有利的方面”，孤独地进行着“一个人的战斗”，并最终取得了成果，他的成果是如此显著，以至于玻尔在战后曾戏称说战时整个欧洲的数学新闻可以归结为一个词，那就是：赛尔伯格。

不过赛尔伯格虽然证明了那个百分比大于零，却并没有在论文中给出具体数值。在赛尔伯格之后，数学家们开始对这一比例的具体数值进行研究，其中以美国数学家列文森（Norman Levinson）的成果最为显著，他证明了至少有34%的零点位于临界线上。列文森取得这一成果是在1974年，那时他已年过花甲，并且行将走到生命的尽头（他第二年就去世了），却依然顽强地从事着数学研究。在列文森之后，这方面的推进变得十分缓慢，几位数学家费尽九牛二虎之力也只能在百分比的第二位数字上做文章，其中包括中国数学家楼世拓与姚琦（他们于1980年证明了至少有35%的零点位于临界线上）。直到1989年，才有人撼动百分比的第一位数字：美国数学家康瑞（Brian Conrey）证明了至少有40%的零点位于临界线上。这也是这方面——并且也是整个黎曼猜想研究中——最强的结果之一，这方面的努力仍在继续。

另外值得一提的是，“黎曼猜想”这一金字招牌后来被推而广之，用来表示一些“山寨版”和“豪华版”的猜想。那些猜想为什么能跟黎曼猜想共享招牌呢？因为它们跟黎曼猜想有极大的相似性，比如都有一个跟黎曼 ζ 函数相类似的函数，那个函数具有与黎曼 ζ 函数相类似的性质，等等。在那些猜想中，“豪华版”黎曼猜想乃是一些比黎曼猜想更强的猜想（即上文提到过的黎曼猜想的推广形式），它们跟黎曼猜想一样，迄今尚未得到证明（这是显然的，否则的话黎曼猜想作为其特例也就被证明了）；“山寨版”黎曼猜想则是跟黎曼猜想有相似性却互不包含的猜想，它们已全部得到了证明，而且撇开我们所取不中听的绰号不论，它们的证明乃是数学上的重大成果，既催生过新数学方法的诞生，也为证明者摘取过数学界的最高奖——菲尔茨奖（Fields medal）。而且，“山寨版”黎曼猜想作为唯一挂着黎曼猜想这一金字招牌却被证明了的猜想，曾使人们对久攻不下的黎曼猜想也一度乐观起来。可惜他山之石，并不总是可以攻玉的。从目前的情况来看，“山寨版”黎曼猜想就只能在“山寨”里玩玩，它们的证明虽然重要，对于解决真正的黎曼猜想却并无实质性的启示。

三

聊了这么多关于黎曼猜想的研究成果，我们稍稍换换口味，来聊一些数学家的故事吧。也许在很多人眼里，数学是一门很枯燥的学问，数学家们则是一群性格乏味的怪人。但实际上，富有智慧的人往往是不会真正乏味的，数学家们也是如此，他们在埋头演算的勤恳之外，也给我们留下了许多独特的幽默。

匈牙利数学家波利亚（George Pólya）曾经讲过一个跟黎曼猜想有关的小故事，故事的主角就是我们前面提到过的英国数学家哈代与丹麦数学家玻尔。这两位对黎曼猜想作出过成果的数学家当然都对黎曼猜想怀有浓厚的兴趣。有一段时间，哈代常常利用假期访问玻尔，一起讨论黎曼猜想，直到假期将尽才匆匆赶回英国。结果有一次，当哈代又必须匆匆赶回英国时，很不幸地发现码头上只剩下一条小船可以乘坐了。从丹麦到英国要跨越几百公里宽的北海（North Sea），在汪洋大海中乘坐小船可不是闹着玩的事情，弄不好就得葬身鱼腹。为了旅途的平安，信奉上帝的乘客们大都忙着祈求上帝的保佑。哈代却是一个坚决不信上帝的人，非但不信，甚至还蓄意跟上帝作对：把向大众证明上帝不存在列入自己某一年的年度心愿之一。不过在那生死攸关的旅程面前哈代也没闲着，他给玻尔发去了一张简短的明信片，上面只写了一句话：“我已经证明了黎曼猜想。”哈代果真证明了黎曼猜想吗？当然不是。他为什么要发这么一张忽悠同事的明信片呢？当他平安抵达英国后他向玻尔解释了原因。他说如果那次他所乘坐的小船果真沉没了的话，那句话就会变得死无对证，人们就只好相信他确实证明了黎曼猜想。可是他知道上帝是绝不会甘心让他这样一个坚决不信上帝的人获得如此巨大的荣誉的，因此它一定不会让小船沉没的。

哈代凭借自己的幽默成为了故事主角，有些数学家则是因为其他数学家的幽默而被动地成为了故事主角，我们前面提到过的法国数学家哈达玛与比利时数学家普森就是如此。这两人成为主角的原因大家恐怕是猜不到的，那是因为他们的长寿：哈达玛享年98岁，普森活到96岁。这两个令人眼红的岁数不知从何时起引发了一个传说，那就是：谁要是能证明黎曼猜想，他就能不朽——不是抽象意义上的不朽（那是毫无疑问的），而是实际意义上的不朽（即长生不老）！不过这个传说的炮制者看来是没有关怀到玻尔和兰道，他们的研究成果可比哈达玛和普森的成果强多了，照说起码也该混个百岁老人当当吧。结果呢？兰道只活了61岁，玻尔稍胜一筹，也只有63岁。可能是意识到这个传说漏洞太大，出生于波兰的数学家欧德里兹科（Andrew Odlyzko）把幽默指向了另一个方向，提出了一个完全相反的说法，那就是：谁要是否证了黎曼猜想，他就会立刻死去！欧德里兹科甚至开玩笑说其实黎曼猜想已经被否证了，只不过那个否证了黎曼猜想的倒霉蛋没来得及发表论文就死去了。

当然，这些都只能作为饭后茶余的谈资而不宜较真。不过，一个极度艰深的东西对投入得过于深入的人产生健康方面的影响，倒并不是毫无可能的。数学界也确实有人猜测，黎曼猜想的极度艰深有可能对个别数学家的健康产生过影响。比如流行传记《美丽心灵》的主角、美国数学家纳什（John Nash）曾在20世纪50年代后期研究过黎曼猜想，在那之后不久就患上了精神分裂症。纳什患病的原因一般认为是参与军方工作引致的心理压力，但也有人认为他贸然去啃黎曼猜想那样的坚果，对其病症发展有可能起到过推波助澜的作用。

黎曼猜想可以说是当今数学界最重要，并且是数学家们最期待解决的数学猜想。美国数学家蒙哥马利（Hugh Montgomery）曾经表示，如果有魔鬼答应让数学家们用自己的灵魂来换取一个数学命题的证明，多

数学家想要换取的将会是黎曼猜想的证明。在探索黎曼猜想的过程中，很多数学家曾经满怀信心，渐渐地却被它的艰深所震动，态度转为了悲观。我们前面提到过的李特伍德就是一个例子，当他还是学生的时候，他的导师就随手把黎曼 ζ 函数写给了他，让他利用暑假时间研究其零点位置。初出茅庐的李特伍德也不当回事地领命而去。后来他与哈代倒也果真在这方面做出了成果。但渐渐地，他的态度发生了变化，甚至表示：“假如我们能够坚定地相信这个猜想是错误的，日子会过得更舒适些。”曾经在“山寨版”黎曼猜想研究上做出过成果的法国数学家韦伊（André Weil）也有过类似的态度转变。当他在“山寨版”黎曼猜想研究上做出成果时，曾经与一些其他人一样对解决黎曼猜想燃起了信心，还表示如果自己证明了黎曼猜想，会故意推迟到猜想提出100周年（即1959年）时才公布——言下之意，自己不迟于1959年就有可能解决黎曼猜想。不过，岁月渐渐磨去了他的乐观，他晚年时曾对一位友人承认，自己有生之年不太可能看到黎曼猜想的解决。就连本文开头提到的那位德国数学大师希尔伯特，他对黎曼猜想的看法也经历了从乐观到悲观的转变。在1919年的一次演讲中，希尔伯特曾表示自己有望见到黎曼猜想的解决，但后来他的态度显著地转为了悲观。据说有人曾经问他：如果他能在500年后重返人间，他最想问的问题是什么？他回答说是：是否已经有人解决了黎曼猜想？

接下来，我们将介绍人们从另一个方向探索黎曼猜想的故事，我们将会看到，那里不仅也有故事，而且还有一些非常出人意料的东西。

四

黎曼那篇提出了黎曼猜想的著名论文除了有许多“证明从略”的地方外，还有一个很突出的特点，那就是它虽然反复涉及了黎曼 ζ 函数的非平凡零点，甚至还提出了与零点分布有关的一系列命题（包括大名鼎鼎的黎曼猜想），却没有举出哪怕一个具体的例子——即没有给出哪怕一个零点的数值。而且与那些“证明从略”的地方并非容易证明同样要命的是，黎曼不曾给出的那些零点的数值也并非轻易就能计算得了的。事实上，直到黎曼那篇论文发表44年后的1903年，才有人填补了这方面的空白：丹麦数学家格兰姆（Gørgen Gram）计算出了15个零点的数值。这是人们首次窥视到黎曼 ζ 函数非平凡零点的具体存在。当然，那15个零点全都位于黎曼猜想所预言的临界线上。

与我们在第二节中介绍的理论研究中的层层推进基本平行，数学家们计算零点的漫长征途，也呈现出了层层推进的态势。但这推进过程在起初一段时间里却显得极为缓慢，直到1925年，才计算出了区区138个零点，而且在那之后陷入了停顿。为什么会陷入停顿呢？原因很简单，就是当时计算零点的方法比较笨拙，致使计算量过于巨大。而当时的计算又全靠手工，零点数目一多，计算量就大到了令人难以应付的程度。

既然是计算方法的笨拙使计算陷入了停顿，那么很显然地，计算的重新启动需要有新的计算方法。这新的计算方法在7年后的1932年终于“出土”了——我没有写错，确实是“出土”，因为它是从早已去世了的黎曼的手稿中“挖”出来的！

黎曼那个时代的一些著名数学家有一个今天的数学家们很少效仿、今天的读者很难理解的特点，那就是常常不发表自己的研究成果。由于

这个特点，那些数学家的手稿有着比普通名人用品所具有的单纯的猎奇价值重要得多的意义，因为从中有可能会发现一些他们未曾发表过的研究成果。黎曼的手稿就是如此。不过令人惋惜的是，黎曼的手稿在他去世后有很大一部分被他的管家付之一炬，只有一小部分被他妻子抢救了出来。在劫后余生的手稿中，又有一部分被他妻子以涉及私人信息为由“克扣”掉了（其中包括许多几乎通篇都是数学，只夹带了极少量私人信息的手稿），剩下的才是后人真正可以查阅的。那些可供查阅的手稿被收录于哥廷根大学的图书馆。

不过，那部分手稿虽然可供查阅，但只要想想黎曼公开发表的文章尚且如此艰深，动辄花费后世数学家几十年的时间才能填补空白，就不难想象研读他的手稿会是什么感觉了。黎曼的研究领域极为宽广，手稿中常常诸般论题混杂，而且几乎没有半句说明。自黎曼的手稿被存放于哥廷根大学图书馆以来，陆续有一些数学家及数学史学家慕名前去研究，但在那极度的艰深晦涩面前，大都满怀希望而来，却两手空空而去。黎曼的手稿就像一本高明的密码本，牢牢守护着这位伟大数学家的思维奥秘。

但到了1932年，终于有一位数学家从黎曼的手稿中获得了重大发现——发现黎曼不仅亲自计算过若干个零点的数值，而且还有自己独特的、直到“出土”之日仍遥遥领先于当时数学界的计算方法。这一发现为黎曼 ζ 函数非平凡零点的计算带来了脱胎换骨般的变化，让停滞在第138个零点上的计算重新启动。当然，这一发现也进一步提高了黎曼那原本就已极为崇高的声望，在很大程度上驱散了一些数学家对黎曼论文中那些“证明从略”部分的怀疑。因为它表明黎曼那篇高度简练的论文只是冰山的尖顶，在那下面有着大量扎实的研究。那么，发现这一切的人是谁呢？是黎曼的一位同胞：德国数学家西格尔（Carl Ludwig Siegel）。为

了从天书般的黎曼手稿中“出土”公式，西格尔付出了艰辛的努力。为了表彰他的努力，人们将这一计算黎曼 ζ 函数非平凡零点的新方法称为黎曼-西格尔公式（Riemann-Siegel formula）。

黎曼-西格尔公式的“出土”大大推进了零点计算。在短短几年间，数学家们就把零点计算推进了一个数量级，达到了1000个以上的零点。虽然随后爆发的第二次世界大战中断了零点计算，但战后计算机技术的发展，又使得零点计算呈现出了井喷势头：从1956年到1969年的十几年间，被计算出的零点数目又推进了好几个数量级，从25000个推进到了3500000（350万）个。当然，所有这些零点也都无一例外地位于黎曼猜想所预言的临界线上。说到这里顺便提醒读者一下，我们这里及下文所说的零点计算除早期那些小规模的计算外，大都只是验证零点是否在临界线上，而并不计算它们的具体数值。

验证了350万个零点全部位于临界线上，无疑大大增强了数学家们对黎曼猜想的信心。不过，不相信的也还是大有人在。比如德国普朗克数学研究所（Max Planck Institute for Mathematics）的一位名叫查基尔（Don Zagier）的数学家对这种验证就不以为然。在他看来，区区350万个零点根本不说明问题。他的这种不以为然很快遇到了对手：一位对黎曼猜想深信不疑的铁杆“粉丝”。这位“粉丝”名叫蓬皮埃利（Enrico Bombieri），是著名的意大利数学家。两人一个疑心重重、一个深信不疑，谁也不服谁。怎么办呢？查基尔提议打赌。说起来，其实查基尔对黎曼猜想倒也并非全然不信，而且也并非一味轻视对零点的数值计算，他只是觉得350万个零点实在太少了，不足以让他信服。那么，要计算多少个零点才能让他信服呢？他开出的数目是3亿个。于是两人就以这个数目为限订下了赌约：如果黎曼猜想在前三亿个零点中出现反例，就算查基尔获胜；反之，如果黎曼猜想被证明，或者虽然没被证明，但在

前3亿个零点中没有出现反例，则算蓬皮埃利获胜。赌注为两瓶葡萄酒。

初看起来，相对于已经计算出的350万个零点来说，查基尔的3亿个零点简直就是“狮子大开口”，查基尔自己也估计这个赌局也许要花上30年的时间才能分出胜负。可是他显然跟那个时代的多数其他人一样，大大低估了计算机技术的发展速度。事实上，离赌局的设立还不到10年，1979年，零点计算就被推进到了8100万个。不久之后，又被推进到了两亿个，距离赌局的终结只剩下了一步之遥，而形势则对查基尔极为不利——因为那两亿个零点全都位于临界线上。

不过，计算出那两亿个零点的数学家对查基尔的赌局一无所知，在计算完两亿个零点后就停了下来，这一点让查基尔大大地松了一口气。可惜，他这口气没能松太久，因为他的一位朋友恰好访问了那位数学家，不仅将赌局之事告诉了后者，还进行了一番鼓动。后者一听零点计算还有这么重大的意义，就立刻展开了新的计算，一鼓作气推进到了3亿个零点——当然，黎曼猜想岿然不动。

查基尔输了，他兑现诺言买来了两瓶葡萄酒。蓬皮埃利当场打开其中一瓶与他共饮。他们喝掉的这瓶葡萄酒用查基尔的话说，是世界上被喝掉的最昂贵的葡萄酒，因为正是为了以它为赌注的那个赌局，数学家们特意多计算了一亿个零点，为此花费了约70万美元的计算经费。也就是说，被他们喝掉的这瓶葡萄酒是用35万美元的经费换来的！喝完了这瓶葡萄酒，查基尔从此也对黎曼猜想深信不疑了。

在查基尔和蓬皮埃利的赌局之后，像查基尔那样看重零点计算、以此决定自己对黎曼猜想信任度的数学家越来越少了；像验证3亿个零点那样愿意把巨额经费投入到零点计算中的人也越来越少了。不过对零点

的计算并没有就此终止。2001年，一位名叫魏德涅夫斯基（Sebastian Wedeniwski）的德国研究者创立了一种崭新的计算模式：分布式计算，即利用彼此联网的许多台计算机来共同计算零点。这个分布式计算系统建成之后，不久就被推向了互联网，吸引了世界各地大量数学和计算机爱好者的参与，联网计算机的数目很快就稳定在了10000台以上，每天计算出的零点数目在10亿个以上。至于经费，则基本可以忽略不计，因为参与者都是自愿而无偿地贡献出自己的计算资源的。

到了2004年末时，魏德涅夫斯基的分布式计算系统所计算出的零点总数逼近了一个激动人心的数目：一万亿个。眼看着一次辉煌庆典已指日可待，不料却从法国传来了一个令人吃惊的消息：两位法国人完成了对10万亿个零点的计算，比他们翘首期待的1万亿个高出了整整一个数量级！更令人吃惊的是，这两位法国人完成这一工作所用的计算资源居然只是几台普通的计算机，所花费的时间也只有一年多。此时此刻，这样的一则消息对于魏德涅夫斯基来说无疑是当头一棒，结果庆典变成了谢幕，魏德涅夫斯基在不久之后关闭了整个系统。此情此景，犹如九十多年前英国探险家斯科特（Robert Falcon Scott）挺进南极的经历：当他们历经艰辛、即将抵达南极点时，却发现挪威探险家阿蒙森（Roald Amundsen）已经捷足先登（斯科特及同伴后来在黯然返回的途中全部遇难）。

两位法国人凭借几台普通计算机一年多的工作，居然超过了全世界上万台联网计算机几年的工作，而且超过了整整一个数量级，这是什么缘故呢？是因为他们采用了一种比黎曼-西格尔公式更高明的计算方法。这一计算方法是出生于波兰的数学家欧德里兹科（Andrew Odlyzko）与合作者肖恩哈格（Arnold Schönhage）于1988年所提出的。

五

欧德里兹科为什么会研究零点计算的算法呢？这也牵扯到一段故事，而且是很有意思的故事。当然，表面上的原因是跟所有其他从事零点计算的人一样的，那就是因为他对零点计算很感兴趣。不过，他那兴趣的由来跟其他人有所不同，其他人的兴趣大都来自于对黎曼猜想本身的兴趣，他却是因为听了美国数学家蒙哥马利（Hugh Montgomery）的一个并非直接针对黎曼猜想的研究报告，才从事零点计算，并研究零点计算的算法的。蒙哥马利那个报告所介绍的是一项很独特的研究，即研究黎曼 ζ 函数非平凡零点在临界线上的分布规律。他的研究表明，在适当的假设——其中包括假设黎曼猜想成立——下，可以证明黎曼 ζ 函数的非平凡零点在临界线上的分布呈现出一种相互排斥的趋势（即倾向于彼此远离），这个趋势可以用一个不太复杂的数学公式来描述。

蒙哥马利自20世纪70年代初就开始研究黎曼 ζ 函数非平凡零点在临界线上的分布规律了。他发现了规律，并且因为那规律不太复杂而直觉地感到在其背后应该蕴含着某种玄机。为了揭开那玄机，他特意访问了普林斯顿高等研究院。在那里，他“觐见”了黎曼猜想研究的元老赛尔伯格。可惜就连赛尔伯格也看不透那规律背后的玄机。不过，在高等研究院那样一个名家云集的地方，随时都有可能出现意想不到的学术交流。蒙哥马利在最有希望得到信息的赛尔伯格那里不曾得到有价值的信息，却在高等研究院的茶室里偶遇了一位物理学家。那位物理学家名叫戴森（Freeman Dyson），是一位研究领域很宽广的人物，当他在和蒙哥马利的攀谈中获知后者所发现的这个零点在临界线上的分布规律时，登时就吃了一惊。因为他想起了自己十多年前的一系列研究。那些研究跟黎曼 ζ 函数的非平凡零点没有半点关系，但在那些研究中，他却得到过同

样的分布规律！

戴森十多年前所研究的是什么呢？是从一些极为复杂的物理体系——比如复杂原子核——中抽象出来的问题。处理那种问题所用的是一类特殊的统计物理手段，而其中一个典型的课题则是研究复杂体系中能量的分布——物理学家们称为能级分布。戴森曾经得到过那种分布的具体形式，它除了可以描述能级外，还出现在了許多其他复杂的物理现象中。而现在，从蒙哥马利所从事的纯数学研究中，他居然再次见到了同样的分布，这实在是大大出乎他意料的事情。

几年之后，蒙哥马利再次来到普林斯顿，并作了一次研究报告——即欧德里兹科所听到的报告。在报告中，他除了介绍自己的研究外，还提到了他和戴森所发现的这种数学与物理之间的奇怪联系。这一切引起了欧德里兹科的浓厚兴趣，使他决定通过大规模零点计算来验证蒙哥马利所发现的零点在临界线上的分布规律。从20世纪80年代末到20世纪90年代初，欧德里兹科利用他和合作者肖恩哈格所提出的新算法，完成了几批大规模的零点计算，结果非常漂亮地证实了蒙哥马利所提出的零点在临界线上的分布规律。考虑到蒙哥马利的结果是在假设黎曼猜想成立的基础上得到的，因此这种证实也可以在一定程度上被视为是对黎曼猜想的间接支持。

不过，所有这些都没有解决一个最根本的问题，那就是像黎曼 ζ 函数非平凡零点在临界线上的分布这样最纯粹的数学性质，为什么会跟像复杂原子核的能级分布那样最现实的物理现象扯上关系？这种神奇的关联本身又预示着什么呢？这两个问题直到今天也没有完全的答案。但有意思的是，在半个多世纪前，却有两位数学家曾经提出过一个猜想——一个与蒙哥马利、戴森、欧德里兹科所发现并证实的这种数学与物理间近乎离奇的联系遥相呼应的猜想。那两位数学家的名字我们在前文中曾

经提到过，一位是希尔伯特，一位是波利亚，那个猜想则被称为希尔伯特-波利亚猜想，它是对黎曼 ζ 函数非平凡零点分布的猜测，其中赫然包括了猜测它们与某个物理体系的能级相对应的可能性！

不过这个希尔伯特-波利亚猜想本身也颇有一些离奇的地方，因为当人们因蒙哥马利、戴森、欧德里兹科的研究而对它发生兴趣，试图追溯它的起源时，却惊讶地发现无论是希尔伯特还是波利亚，居然都不曾在任何文字之中述及过这个猜想。难道这个猜想根本就是子虚乌有的传说？幸运的是，94岁高龄的当事人波利亚那时仍健在，他在一封信件中以个人回忆的方式肯定了这一猜想的存在性。但早已去世的希尔伯特在什么场合下提出过这一猜想，却很可能将成为数学史上一个永久的谜团了。

六

介绍了这许多有关黎曼猜想的研究，有一个问题想必很多读者都会关心，那就是黎曼猜想的终极命运将会如何？它是会被证明呢？还是会被推翻（否证）？对于这个有关黎曼猜想“前途命运”的大悬念，数学家们各有各的看法。

有些数学家相信黎曼猜想是对的，比如那位输掉了葡萄酒的查基尔自赌局告负之后就对黎曼猜想深信不疑。他相信黎曼猜想的理由很“纯朴”，那就是数值证据已经够强了。读者们想必还记得，他当时要求的数值证据是3亿个零点，现在的证据已经超过了10万亿个，远远超出了他的要求。因此，他的相信是有理由的。不过，由于零点有无穷多个，实际上再多的数值证据也是微不足道的。而且在数学上有过这样的例子，即一个被否证了的数学命题的数值反例出现在极遥远的地方，远远超出数值证据所能触及的范围。黎曼猜想会不会也是如此呢？谁也说不准。当然，支持黎曼猜想的证据不仅仅来自数值计算，还有我们介绍过的大量其他研究，其中包括至少有40%的非平凡零点位于临界线上那样颇为可观的结果。相信黎曼猜想的数学家们也可以从那些方面获得信心。

有些数学家则认为黎曼猜想是错的。面对黎曼猜想所得到的如此海量的支持，选择那样的立场当然是要理由的。这其中一条打不倒的理由就是：所有支持都不是证明。确实，对于像黎曼猜想这样的数学命题来说，要想证明它成立，必须“一个都不能少”地涵盖所有的零点，缺一丁点儿都不行。但反过来，要想推翻它，却只要找到一个反例——即一个不在临界线上的非平凡零点——就足够了，这种繁简程度上的不对称对于怀疑黎曼猜想的数学家们是十分有利的。

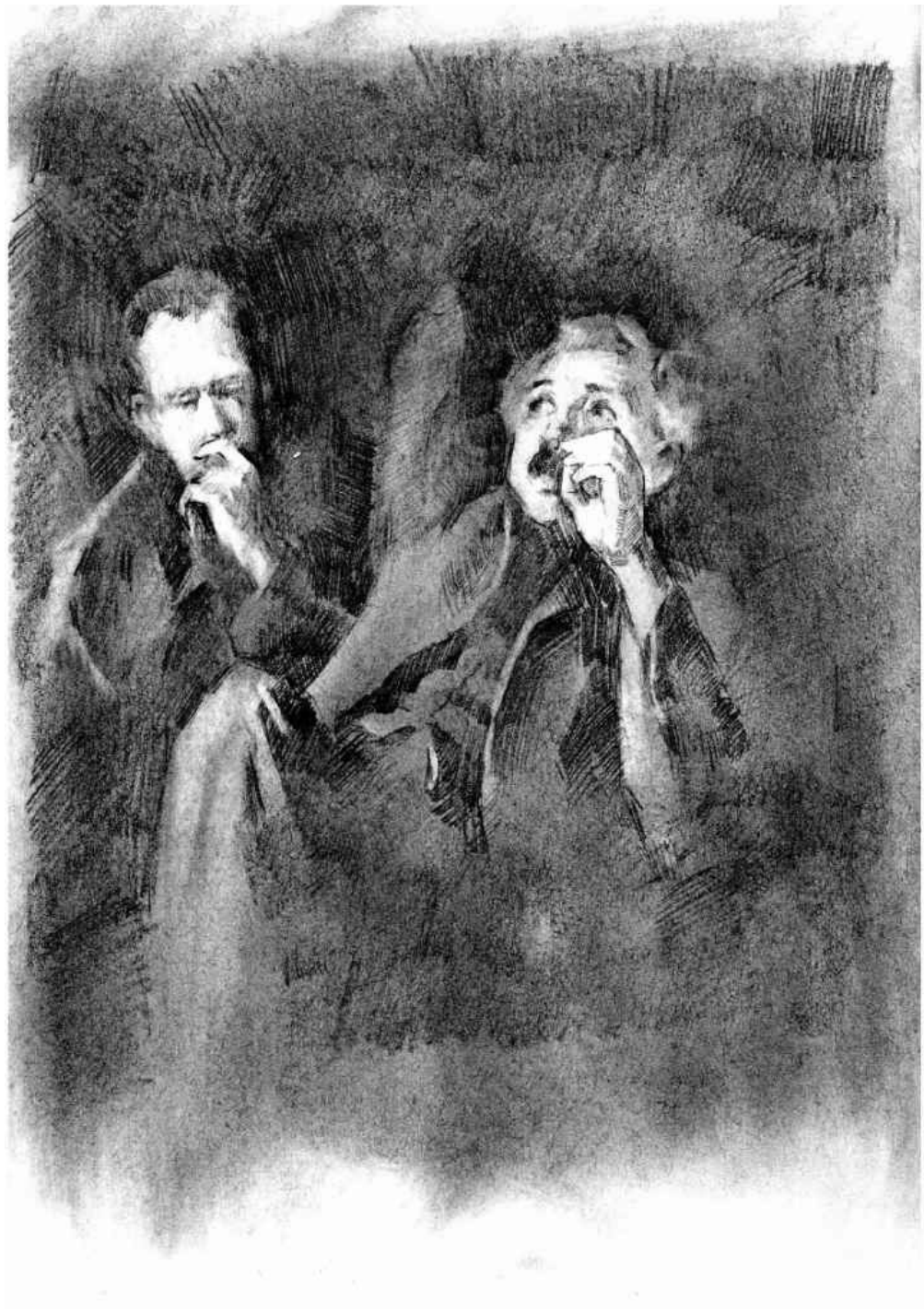
除上述两种截然相对的态度外，黎曼猜想的长期悬而未决还使一些人联想到了所谓的哥德尔不完全性定理（Gödel's incompleteness theorem），认为黎曼猜想有可能是一个不能被判定——即既不能被证明，也不能被否定——的命题。据说哥德尔（Kurt Gödel）本人就有过这样的看法。不过，黎曼猜想假如不成立，在原则上是可以用明确的步骤，通过数值计算找到它的反例，从而证明其是不成立的。从这个意义上讲，黎曼猜想假如不成立，它是可以被判定为不成立的，而它如果不能被判定，实际上是表明它成立。

好了，以上就是对黎曼猜想的简单介绍。这一介绍因为略去了数学细节而看上去更像是一串故事。但实际上，黎曼猜想是一个极为艰深的课题，如果哪位读者想要啃一啃这个猜想，首先要有扎实的数学功底，否则非但啃不动，还很可能会崩掉牙齿——可别怪我没提醒哦。

二零一二年三月四日写于纽约

⁽¹⁾ 本文曾发表于2012年3月8日和15日的《南方周末》，发表稿经编辑修改，发表时的标题分别为《素数之魂——黎曼和他的伟大猜想》和《十万亿个证据不如一个证明——猜猜黎曼猜想的命运》。

第二部分 其他



科学的目的

The most incomprehensible thing about the world is that it is comprehensible.

—Albert Einstein

很佩服做哲学研究的人，无论什么话题都能够洋洋洒洒地写出几十万言。有时候觉得他们有点像诗人，小中见大，平中见奇，能够把一个简单的概念写复杂了。一部沉甸甸的著作常常会引起人们本能的敬畏，一句听起来似懂非懂的话常常让人觉得“嗯，有点哲学味”。这种敬畏，这种“哲学味”，在一定程度上使大众疏远了科学。曲高则和寡，自古如此。

科学的数学结构是抽象的，但科学的理念却是朴素的。⁽¹⁾当爱因斯坦为指南针神秘的方向性感到惊讶时，他只有4岁，还没有来得及读亚里士多德，也还看不懂康德。后来人们认为爱因斯坦是个天才，但那时候的他还只是一个晚熟的孩子。他感到了惊讶，因为他不知道事情为什么会这样；他后来成为了物理学家，因为他想知道事情为什么会这样。只有真正朴素的理念才能和一个4岁孩童的朦胧理性产生耦合，而我深信一个真正朴素的理念是不需要用几十万字才能说清楚的。

这个宇宙的演化是有逻辑规律的，这个宇宙间纷繁多姿的现象背后是有原因的，这是科学存在的前提，也是任何智慧存在的前提。至于这个宇宙为什么是有逻辑规律的，这并不属于科学的范畴。我们存在于这

样一个宇宙之中，这是一个基本的经验事实。^②这个经验事实也意味着逻辑推理的有效性是一个近乎于先验的基本事实。

就像宇宙间所有的其他存在一样，科学的存在也是有因果的，科学存在的具体形式是和它所要达到的目的紧密相联的。环顾我们周围的世界，从草木竹石到飞禽走兽，从戈壁草原到冰川湖泊，小至蝼蚁尘埃，大至日月星辰，世间的现象是如此的千变万化，无穷无尽，就像满地的珍珠，如若没有丝线相串，何以尽拾？科学也是这个道理，万物无穷而人力有限，理解事物的唯一有效的方法就是简化。把许多现象归结为一个道理，窥一隅而知全貌，就是一种最有效的简化。寻求对自然现象的这种简化是人类试图理解、预言和利用自然现象的最重要途径，也是科学朴素而优美的目标。

当然，我们也应当看到，“简化”是一个比较含糊的字眼，不问内涵地追求简化会使人误入歧途。最大而又最荒唐的简化莫过于把一切归因于上帝，就像《圣经》所宣称的，那比牛顿定律、麦克斯韦方程式，或相对论的基本原理简单多了。但那不是科学，因为《圣经》只不过是把它所要“解释”的东西罗列了一遍，上帝第一天创造什么，第二天创造什么……如此而已。哪怕略去其中无数的错误不论，这种所谓的“解释”除生添一个上帝外，也并不构成任何实质意义上的简化。更重要的是，这种“简化”缺乏人们对科学的一个很基本的期盼，那就是要能够预言未知或未来的现象。仅限于对已知及已经发生过的现象进行罗列、归纳或整理，哪怕做得很到位，也更接近于历史而不是科学。

那么，对自然现象什么样的理解能够构成实质意义上的简化，并且具有科学所必须具有的预言能力呢？是以逻辑推理为依据的理解。把科学的理论框架建立在逻辑推理之上是其力量的重要源泉，也是科学有别于宗教的一个极其本质的特征。在一个科学理论中，从基本原理到对现

象的解释，是以逻辑推理的方式来衔接的。由于——如前所述——逻辑推理的有效性是一个近乎于先验的基本事实，我相信人类远在意识到“逻辑”这个概念之前，就已经在本能地运用着初等的逻辑推理了。逻辑推理具有极大的延展性和客观性。从一个科学理论的基本假设出发，运用逻辑推理可以衍生出近乎于无限的推论，而且这些推论是以非常确凿并且独立于个人意志的方式存在着的。一个科学理论一旦提出，就以一种严谨而谦虚的方式存在于学术界。任何人都有权对它的基本假定和逻辑推论进行检验。任何一个那样的检验如果得出明确的否定结果，就意味着理论被推翻，或者其局限性被发现。科学理论的这一特征被科学哲学家波普尔（Karl Popper, 1902—1994年）提升到了一个很核心的地位。波普尔写过许多大部头的书，其中一个基本的观点，就是认为一个理论成为科学理论的必要条件是这个理论具有可证伪性

（falsifiability）。也就是说一个理论要成为科学理论，必须明确地提出在何种情形下自己可以被推翻。这一点初看起来很出人意料，因为通常人们在思考科学理论时，往往是从证明而不是证伪的角度去考虑的。但细想一下其实却不难理解，因为一个科学理论的推论是无穷尽的，再多的实验也只能加强它的可信性而无法证明它的正确性。相反，由于科学理论有着明晰的逻辑结构，要推翻它却只要有一个确凿的反例就可以了。

人性有弱点，科学家是人，因而也不例外。疏忽、偏见，甚至蓄意的伪造都有可能带来谬误。科学之所以能够在探索自然的漫长征途中去芜存菁，获得卓越的发展，正是得益于科学理论严密的逻辑性和科学界这种公正、谦虚和理性的态度，这是人类智慧的骄傲。^③

综上所述，科学的目的可以大致地叙述为：科学寻求的是对自然现象逻辑上最简单的描述。

二零零二年五月二十三日写于纽约

(1) 本文所说的科学是指自然科学。

(2) 喜欢“人择原理”的话，可以认为假如这个宇宙不是这样的，那就不会有任何“人”来问“宇宙为什么会是这样的”。这不仅是因为在一个没有逻辑规律的宇宙中不可能产生所谓的智慧生命（想一想什么是智慧），而且也是因为问问题本身就是一种逻辑的思维方式。在一个没有逻辑规律的宇宙中，这样的思维方式是没有意义的。因此我们在这里问这个问题本身就已经假定了宇宙是有逻辑规律的。

(3) 相形之下，政治舞台上不容挑战的“伟大、光荣、正确”，宗教神坛上不容挑战的“唯一的真神”之类的自我标榜是何等的虚伪。

科学的方法

The most incomprehensible thing about the world is that it is comprehensible.

—Albert Einstein

我撰写的最早的科学哲学短文是《科学的目的》，写于2002年。那篇短文当初的标题是《什么是科学（一）》，本拟写成一个系列。后来由于未曾续写，便改名为《科学的目的》，与其他几篇短文一起并入了一个合集《什么是科学》。不过，那个合集虽取了当初那个系列的标题，其实却只是不同时期、不同场合所写的几篇科学哲学类短文的松散组合，而非原先拟写的那个系列。

原先那个系列一搁置便是7年。2009年，我将自己一些旧作的合集制作成了PDF文件，其中包括了那个科学哲学短文的合集。那个合集PDF的制作使我重新考虑起了当年那个系列。虽然这些年来我对科学哲学的评价越来越低，写作兴趣越来越小，不过当年拟定的写作思路我觉得仍是有意义的。那个思路简言之就是先谈目的，后论方法。之所以设定那样的思路，是因为方法往往取决于目的，目的一旦确定了，讨论方法就有了参照。我曾在不止一篇文章中提到过科学方法优于其他认知方法，理由何在呢？在本文中，我就重拾昔日的思路，以科学的目的为参照，来谈谈这一问题。

我们首先重复一下《科学的目的》一文中介绍过的科学的目的：

科学的目的是：科学寻求的是对自然现象逻辑上最简单的描述。

这一目的对于科学来说几乎是定义性的。虽然科学也被人们用来谋求很多其他的目的——善良的或邪恶的，政治的或军事的，社会的或文化的——但那些只是科学的应用，以及某些科学家从事科学的动机，而非科学本身的目的。

除科学的目的外，在本文中我们还将用到一个有关科学的基本事实，那就是：

有关科学的基本事实：我们并无任何已被确认的、能理解全部自然现象的科学理论。

上述事实应该是足够显而易见的，而且显然是被科学界所普遍认可的。

明确了科学的目的及上述有关科学的基本事实，我们就可以以之为出发点，来回答一个对本文来说具有核心意义的问题，即什么样的方法是我们追求科学的目的时应该采用的正确方法？这个问题的答案并不复杂。我们首先注意到：既然我们并无任何已被确认的、能理解全部自然现象的理论，那么为追求科学的目的所提出的任何东西就都存在出错的可能性。而既然存在出错的可能性，那么纠错就是必不可少的。因此，追求科学目的的正确方法所须满足的第一个基本特征就是允许纠

错，并且具有纠错能力：

追求科学目的的正确方法所须满足的基本特征之一：允许纠错，并且具有纠错能力。

既然需要纠错，那么接下来的一个很自然的问题就是：以什么为依据来纠错？这个问题的答案也很简单：既然科学寻求的是对自然现象逻辑上最简单的描述，那么纠错的依据显然就是自然现象及逻辑推理。由于我们了解自然现象的基本途径是观测与实验，[\(1\)](#)因此追求科学目的的正确方法就必须尊重观测与实验，尊重逻辑推理。这是纠错的依据，也是追求科学目的的正确方法所须满足的第二个基本特征：

追求科学目的的正确方法所须满足的基本特征之二：尊重观测与实验，尊重逻辑推理。

这两个基本特征正是科学方法的基本特征，反过来说，任何方法只要切实满足上述基本特征，就是科学方法，这可以说是科学方法的定义（或定义的一部分）。虽然科学哲学的推理大都是模糊的，有时甚至是似是而非的，但上述推理在我看来是例外，只要认同前面所列的科学的目的及基本事实，上述特征就是必然推论，在这点上并无多少模糊性。从这个意义上讲，科学方法不仅优于其他认知方法，而且还是追求科学目的的唯一正确的方法。

当然，这里有必要说明一点，那就是不用科学方法，或者说用不正

确的方法，未必一定不能产生出对科学有用的东西，它只是不能系统性地产生那样的东西，而且随着科学的不断发展，不用科学方法而产生出对科学有用的东西的可能性会越来越小。为民科辩护的人常常会说：民科的东西未必一定是错误的。说得很对，其实别说是民科的东西，哪怕是胡乱敲击键盘的猴子敲出的东西，也会有一个不为零的概率能包含一些正确的东西（是否有人愿意为了那样的概率来阅读那样的东西，就另当别论了）。⁽²⁾但是，不管用什么方法得到的东西，如果要成为科学理论，就必须接受科学方法的检验。⁽³⁾

在本文的最后，我们稍稍扯远一点。参与或旁观过有关科学与宗教的讨论的读者也许大都见过这样的场景，那就是每当有人提出科学优于宗教时，往往就会有人反驳说科学也有不能解释的东西。⁽⁴⁾在这里我们要顺便强调一点，那就是：我们支持科学，首先支持的是科学的方法，其次才是科学的内容。科学与人类对未知世界的其他一些认知方式之间最本质的差异，在于他们对待未知及对待自身的态度迥然不同。科学的自我纠错、自我完善的能力是许多其他认知方式所不具有的。许多人以科学也有不能解释的东西来为宗教等辩护，看似思路开阔，不拘泥于已有的知识体系，实际上却忽略了一个最重要的地方：那就是正因为已有的知识体系存在局限性，才更需要一种像科学那样能够客观理性地对待自身局限性，随时愿意接受证伪的认知方式，这是我们认为科学远远优于其他认知方式的最深层的原因。正是因为科学具有这种素质，我们通过科学方法所获得的知识才远比通过其他认知方式所得到的东西更客观，更接近正确，也更有希望达到正确——如果“正确”是可以达到的话。但这种通过科学方法所获得的知识——即科学的内容——尽管已是人类知识中最值得珍视的部分，它的价值与科学方法相比仍是其次的。只要科学方法存在，哪怕所有具体的知识都失去了，使科学的直接解释能力暂时降到与宗教一样薄弱的地步，假以时日，科学依然能够重新发

展起来，并远远超越其他认知方式。

二零零九年六月五日写于纽约

(1) 我通常用“观测”表示对来自于自然现象的信息的直接接收（比如观测行星的运动），而用“实验”表示对人为安排过的现象所做的观测。不过实际上，哪怕“观测”也并不是完全直接的，而是与理论有关（比如任何通过视觉的观测都依赖于光学理论）。观测和实验与理论之间的关系实际上是一种相当复杂的依赖关系。不过它们之间必不可少的自洽要求还是能让我们区分好的与坏的理论体系。比方说，如果一个理论体系认为真空中的光线是七扭八歪的，那么我们观测到的椭圆轨道经过这种光学理论一反推，就会对应于相当复杂的行星轨道。相应的，引力理论就不可能如牛顿引力那样简单。那样的理论体系——如果能够建立起来的话——就会明显不如我们现有的理论体系来得优越。

(2) 需要说明的是，在这里提到胡乱敲击键盘的猴子并非是要用一个概率更小的例子来贬低民科。事实上，从文字上碰对一些东西的概率，固然是民科远远高于胡乱敲击键盘的猴子；但另一方面，一涉及具有一定复杂度的技术性内容，情况就恰好相反：胡乱敲击键盘的猴子会有一个不为零的概率碰对，而民科由于翻来覆去只会在自己那口井里扑腾，碰对的概率反倒是零。因此这两者是互有长短，而非一者贬低另一者的关系。

(3) 当然，如前所述，方法往往取决于目的，科学的方法是为了追求科学的目的而采用的方法，在所有可能的方法中，它只是一个小小的部分。在这个世界上，不同的人有不同的追求，如果所追求的不是科学的目的，那么相应地，所采用的也就不一定是——甚至一定不能是——科学的方法。比方说，如果所追求的是某种不容挑战的教条——比如《圣经》或某某主义，那么所采用的就会是一些截然不同的方法，比如自欺、欺骗、附会、偷换概念、混淆逻辑、答非所问、选择性目盲、选择性遗忘、请不听话的人“喝茶”、请很不听话的人上“火刑柱”，等等，这些都不是科学的方法。

(4) 有关这一点的进一步讨论，可参阅拙作《科学哲学讨论中的“大规模杀伤武器”》（已收入本书）。

科学哲学讨论中的“大规模杀伤武器”

在科学哲学讨论中，常常有人提到这样一种观点：即科学不是万能的，科学也有出错的时候。说实话，我很怕在普通讨论中看到这种观点。为什么怕呢？倒不是因为这种观点不正确，相反，正是因为它非常正确。我们现在甚至未来所具有的任何知识（包括科学知识）都不仅未必是终极真理，而且几乎一定不会是终极真理。哪怕有一天我们真的发现了终极真理，也不会有办法证明那就是终极真理。因此这个观点正确得无以复加。事实上，这个观点是如此正确，它几乎是科学哲学讨论中的“大规模杀伤武器”（weapon of mass destruction）。

如果在科学哲学讨论中，有人糊涂到以为我们的知识是绝对正确的，那这种“大规模杀伤武器”是很有必要拿出来用一用的。但在一般讨论中，动辄祭出这类法宝却最容易伤及无辜，产生混淆视听的作用，这远比提出一个错误观点有害得多。后者只要反驳就行，有时非但无害，反而有助于澄清观点；而前者却往往驳也不是，不驳也不是。这有点像现实生活中的好心办坏事，你责备也不是，不责备也不是，比真正的坏事还棘手。

比如在讨论科学与宗教作为认知方式的差异与优劣时，就常常有人提出这种“科学不是万能的，科学也会出错”的观点。它显然很正确。但在那种讨论中，在没有人声称科学永远正确的情况下，主动插入这种观点，却明显起到了用所有认知方式都非完美无缺这一特点，来混淆不同认知方式的优劣之分的作用。这是极具误导性的。

如果把话题稍稍延伸一点的话，那么类似地，在讨论学习方法，或向年轻网友建议学习方法时，也存在这样一种杀手锏：那就是历史上几

乎所有的学习方法，科班也好，自学也罢，甚至连民科都算上，往往都可以举出个别成功的例子。因此每当有人主张或建议某种学习方法时，就往往会有人提出反例，表明未必需要遵循这种学习方法。比如在讨论民科问题时，印度数学家拉马努金（Srinivasa Ramanujan）就是一个常被人提出的例子。虽然拉马努金在英国受过哈代（Godfrey Hardy）的训练，从而明显有别于普通民科，但他到达英国之前，在远离学术界的情况下凭一己之力获得的成果也颇为可观。这是否意味着他的学习方法与其他学习方法有了同等的推荐价值呢？显然不是。因为拉马努金所具有的奇异数学天赋别说普通人，即便在数学界的顶尖人物之中也是无人可及的。^①当我们向他人建议学习方法时，与其假定对方具有拉马努金的资质，不如假定对方是常人（除非已经有迹象表明对方不是常人，即便如此，对方资质接近拉马努金的可能性依然微乎其微），这样的假定更有可能接近事实，所提的建议也更可能有价值。

当然，这里我们要对建议与规定做个区分。如果我们讨论的是规定，即对方必须照办的东西，则应该采取相反的思路，即不排除一个人的资质接近拉马努金的可能性，给人留出一定的机会来证明自己究竟是不是拉马努金。只有这样，才能做到不扼杀拉马努金式的天才。那样的天才虽然极其罕见，但也极其珍贵，在制度层面上应该为他们预留机会。但建议则不同，简单地讲，我主张规定从宽，建议从严。建议的东西采纳与否全在对方，对方甚至可以一方面接受，一方面仍按自己的方式去做，直到失败了再回过头来尝试建议。因此在建议中——除非已经有迹象表明对方不是常人——我们告诉对方的往往是我们认为对常人来说最有可能成功的方式，而不是对无论多小的可能性都给予同样的重视（否则等于什么也没建议）。如果我们在向别人建议的时候，预先假定对方是拉马努金式的天才，无需接受系统的知识就可以闭门造车，那么我们的建议或许有0.00000001%的可能性是鼓励了一个真正拉马努金式

的天才，却有99.99999999%的可能性是在误导一个普通人（甚至误导一个具有科学家潜质的人才）。

我们都知道，在现实世界里几乎没有什么东西是绝对的。我们讨论问题的时候，往往不可能周密到每一句话都体现出这种非绝对性。在许多情况下，我们指的只是可能性较大的东西。假如我们向对方所提的每一个建议都要兼顾对方资质接近拉马努金这种可能性，那么我们或许也应该在每一句话中加上诸如“倘若你没有在街上被汽车撞到”之类的假定，因为那也是会影响对方成材的可能性，而且其概率恐怕比对方是拉马努金的概率大得多。如果大家一发现谁的某句话存在遗漏的可能性，就拿出那种“大规模杀伤武器”来对付，则往往看似正确，其实却使讨论误入歧途，就像在真正的大规模杀伤武器作用下，好人和坏蛋一起丧命一样。

如果我们真想把话说得滴水不漏，最好的办法或许是在每一段话的开头都先附上类似于罗素的“自由思想十诫”之类的东西，然后在“但是，……”之后表述自己的观点。

二零零六年三月二十五日写于纽约

⁽¹⁾ 本文在我网站发表后有网友对这一说法提出了异议，我在这里补充说明几点：第一，如上文所述，这里讨论的是拉马努金具有的那种奇异数学天赋，而非对数学的贡献或影响力，论后者许多数学大师远在拉马努金之上。第二，这里讨论的是天赋的一种，即拉马努金对数字与公式的近乎神奇的判断力，而非其他类型的天赋，论后者许多数学大师远在拉马努金之上。第三，拉马努金的天赋与其他类型的天赋（比如超强的领悟力）究竟哪一种对数学家更为重要，不是本文讨论的内容。第四，其他数学家的天赋或是体现在比他人更早具备某种能力（比如阅读某些艰深著作的能力），或是体现在比他人具有程度更高的某种能力。这些能力虽然令人敬佩，但只是渐变式的优势，不会让我觉得不可思议。拉马努金的天赋却是那种常人（包括数学家）无论训练多少年都难以企及的，是一种“飞跃”。在我看来，拉马努金是最接近“天才”一词

字面意义的数学家。

关于批评的资格

除了上文介绍的“大规模杀伤武器”外，还有一种“武器”在讨论中也具有很大的杀伤力，而且使用范围不限于科学哲学讨论。如果要比喻的话，或许可以比作美国军方研制的威力仅次于核武器的所谓“炸弹之母”。这种“武器”就是：如果你未曾深入了解某种东西，你就没有资格去批评。

粗看起来，这个观点也是很正确的，就像某革命导师所说的“没有调查研究，就没有发言权”。但是正如该导师的很多言论曾让很多人吃足苦头一样，这个很具杀伤力的“炸弹之母”也很容易对讨论产生混淆作用。而且这个“炸弹之母”还有一个厉害之处，那就是它针对的不是观点而是资格，一旦剥夺了资格，就无需再对观点进行反驳了，因而特别适用于对付那些用其他手段难以对付的观点。而且更糟糕的是，这种“炸弹之母”只要稍加混淆就会有很广泛的“适用”范围，从而获得额外的杀伤力，因为在很多情况下，被批评的东西往往是那些在批评者看来错误或没有价值的东西，而恰恰对那样的东西批评者是不会化大力气去“深入了解”的（因为那将是浪费时间）。这样一来，就彻底地进入了“炸弹之母”的攻击范围。更何况，就算批评者对批评对象有过“深入了解”，这“深入”的标准及“资格”的门槛都是可以无限抬高的。对所有并非爱因斯坦的人来说，这“资格”二字都有可能变为不可承受之重（其实就连爱因斯坦也曾在自己专业之外的政治、宗教、文化等领域发表过大量的批评或评论，从而也未必能完全不受“炸弹之母”的威胁）。

拿我自己作个例子，我并非宗教专家，却写了一篇《论宗教》，⁽¹⁾这就也有了一个资格问题，即我有没有资格写那篇文章？我的看法是：谈论“深入了解”或批评的“资格”必须以所涉及的具体论点为参照。拿

我的《论宗教》来说，它并不是一篇泛评宗教的文章，而是只针对宗教活动中用科学附会教义、信徒对教义的盲信等几种特殊现象提出批评。宗教在这些方面的活动及特点是很公开的，并且有大量的传教宣传品为证。作为在美国这样一个基督教盛行的国家生活了十几年的人，我有很多机会接触到教徒，接收到传教宣传品，并与试图向我传教的人进行辩论。这些虽不足以使我成为宗教专家，却足以确立我所批评的那些现象的确凿存在。而只要那些现象确凿存在，我的批评就具备了事实依据，我对批评对象的了解也就达到了支持文章论点所需的深入程度。如果说资格的话，这就是资格，它是以我的具体论点为参照的。至于我是否通晓《圣经》的每一页写了什么东西，我是否是宗教专业毕业的，那根本就与我的论点无关，从而不能与我写那篇文章的资格混为一谈。

像这种针对社会现象而提出的批评或评论是极其常见的，几乎每位作者——无论他是博客、论坛还是传统媒体的作者，也无论其专业是什么——都多多少少会写一些。这类文章的观点正误当然可以争论，但只要所针对的社会现象确凿存在，作者的资格就不成问题。接下来要看的是各人对所针对的社会现象进行分析的能力，而这种能力往往是各专业的作者都具有的，而且也未必是那种“术业有专攻”的文科或社科专业的作者最强（因为分析能力往往是他们的弱项）。当然，这里会有例外，比如涉及法律的东西，除了现象确凿，具有分析能力之外，了解法律法规也很重要，在这类问题上专业作者具有特殊优势。

上面这些论述，并不是要否认批评资格问题的存在。事实上，批评者具备批评资格的情况固然很多，批评者缺乏批评资格的情况也并不少见。比方说，很多民科宣称自己推翻了相对论。让我们用同样的标准——即谈论“深入了解”或批评的“资格”必须以所涉及的具体论点为参照——来看看他们的批评资格。他们所涉及的具体论点是什么呢？是推翻

相对论。事实依据是否确凿呢？不确凿，因为迄今并不存在与相对论确凿矛盾的实验。我们再看他们的分析能力，与多数社会现象不同，对相对论那样的现代物理理论进行分析，除了要有基本的逻辑分析能力，还要有相当程度的数学演算能力，以及对某些实验结果的含义的充分了解（在需要额外知识或能力这点上，它类似于前一段末尾提到的涉及法律这一例外情形）。一个批评者如果是那种一无事实论据，二无数学能力，三无实验基本常识的“三无”人员，却试图推翻被众多定量预言与精密实验所确立的物理理论，他的资格当然就成问题了。这时候——也只有在这类时候——人们才应该用“炸弹之母”来对付。

最后需要说明的是，具备了批评资格当然并不意味着批评的内容正确，对后者大家可以各抒己见，但讨论的主要依据应该是逻辑与证据，而不是动辄使用“大规模杀伤武器”或“炸弹之母”。

二零零九年六月八日写于纽约

(1) 对此文感兴趣的读者可到我网站 (<http://www.changhai.org/>) 浏览。

人名索引

阿哈罗诺夫 (Yakir Aharonov)

阿蒙森 (Roald Amundsen)

艾伯纳 (Dieter Ebner)

艾伦菲斯特 (Paul Ehrenfest)

艾萨克森 (Walter Isaacson)

爱丁顿 (Arthur Eddington)

爱因斯坦 (Albert Einstein)

奥本海默 (J. Robert Oppenheimer)

奥格·玻尔 (Aage Bohr)

奥斯特瓦尔德 (Friedrich Ostwald)

贝尔 (John Stewart Bell)

贝索 (Michael Besso)

毕安基 (Luigi Bianchi)

毕尔克尼斯 (Christopher Bjercknes)

波利亚 (George Pólya)

波普尔 (Karl Popper)

波斯特 (Emil Post)

玻恩 (Max Born)

玻尔 (Harald Bohr)

玻尔 (Niels Bohr)

玻姆 (David Bohm)

玻特 (Walther Bothe)

仓格尔 (Heinrich Zangger)

查基尔 (Don Zagier)

戴森 (Freeman Dyson)

戴维斯 (Martin Davis)

德拜 (Peter Debye)

德布罗意 (de Broglie)

德斯帕那特 (Bernard d'Espagnat)

狄拉克 (Paul Dirac)

丁肇中

丢番图 (Diophantus)

厄尔曼 (John Earman)

恩兹 (Charles Enz)

法蒂夫 (D. K. Faddeev)

费恩曼 (Richard Feynman)

费马 (Pierre de Fermat)

费米 (Enrico Fermi)

芬克尔斯坦 (David Finkelstein)

冯·弗莱克 (John Hasbrouck Van Vleck)

弗兰克 (James Franck)

弗伦奇 (Anthony French)

福克 (Vladimir Fock)

福勒 (Ralph H. Fowler)

伽莫夫 (George Gamow)

盖革 (Hans Geiger)

戈革

哥德尔 (Kurt Gödel)

格尔克 (Ernst Gehrcke)

格兰姆 (Gørgen Gram)

格里莫尔 (Clark Glymour)

格罗斯曼 (Marcel Grossmann)

宫苏艺

郭汉英

哈达玛 (Jacques Hadamard)

哈代 (Godfrey Hardy)

海森伯 (Werner Heisenberg)

赫尔布兰德 (Jacques Herbrand)

胡大年

惠勒 (John Archibald Wheeler)

惠塔克 (Andrew Whitaker)

惠特克 (Edmund Whittaker)

霍尔顿 (Gerald Holton)

霍夫丁 (Harald Høffding)

卡拉西奥多里 (Constantin Carathéodory)

卡鲁查 (Theodor Kaluza)

康普顿 (Arthur Compton)

康瑞 (Brian Conrey)

科里 (Leo Corry)

克尔凯郭尔 (Søren Kierkegaard)

克拉默斯 (Hendrik Anthony Kramers)

克莱因 (Oskar Klein)

Kronig (Ralph Kronig)

克林 (Stephen Kleene)

库尔恰托夫 (Igor Kurchatov)

拉马努金 (Srinivasa Ramanujan)

兰道 (Edmund Landau)

朗道 (Lev Landau)

劳厄 (Max von Laue)

勒波维茨 (Joel Lebowitz)

勒纳德 (Philipp Lenard)

雷恩 (Jürgen Renn)

黎曼 (Bernhard Riemann)

李特伍德 (John Littlewood)

李政道

里奇 (Gregorio Ricci-Curbastro)

栗弗席茨 (Vladimir Lifshits)

列文森 (Norman Levinson)

吕丁格 (Erik Rüdinger)

楼世拓

卢瑟福 (Ernest Rutherford)

鲁本斯 (Heinrich Rubens)

罗巴切夫斯基 (Nikolai Lobachevsky)

罗宾逊夫人 (Julia Robinson)

罗伯森 (Peter Robertson)

罗瑟 (John Rosser)

罗森菲尔德 (Léon Rosenfeld)

洛伦兹 (Hendrik Lorentz)

马蒂亚塞维奇 (Yuri Matiyasevich)

马尔科夫 (A. A. Markov)

马赫 (Ernst Mach)

马斯洛夫 (S. Yu. Maslov)

麦克斯韦 (James Clerk Maxwell)

梅拉 (Jagdish Mehra)

梅西亚 (Albert Messiah)

蒙哥马利 (Hugh Montgomery)

米 (Gustav Mie)

米尔斯 (Robert Mills)

米列娃·爱因斯坦 (Mileva Marić)

穆尔 (Ruth Moore)

纳什 (John Nash)

南部阳一郎 (Yoichiro Nambu)

能斯特 (Walther Nernst)

牛顿 (Isaac Newton)

诺特 (Emmy Noether)

欧德里兹科 (Andrew Odlyzko)

派尔斯 (Rudolf Peierls)

派斯 (Abraham Pais)

潘诺夫斯基 (Erwin Panofsky)

泡利 (Wolfgang Pauli)

彭罗斯 (Roger Penrose)

蓬皮埃利 (Enrico Bombieri)

普朗克 (Max Planck)

普森 (Charles de la Vallée-Poussin)

普特南 (Hilary Putnam)

丘奇 (Alonzo Church)

屈傲诚

荣格 (Carl Jung)

赛尔伯格 (Atle Selberg)

施塔赫尔 (John Stachel)

束星北

斯科特 (Robert Falcon Scott)

斯莱特 (John Slater)

斯塔克 (Johannes Stark)

斯陶耳岑堡 (Klaus Stolzenburg)

斯特恩 (Otto Stern)

斯特劳斯 (Ernst Straus)

苏步青

苏士侃 (Leonard Susskind)

索波列夫 (Sergei Sobolev)

索尔 (Tilman Sauer)

索末菲 (Arnold Sommerfeld)

塔尔斯基 (Alfred Tarski)

泰勒 (Edward Teller)

汤川秀树 (Hideki Yukawa)

图灵 (Alan Turing)

托德洛夫 (Ivan Todorov)

托梅尔普劳兹（Senta Trömel-Plötz）

外尔（Hermann Weyl）

王淦昌

韦斯科夫（Victor Weisskopf）

韦伊（André Weil）

维格纳（Eugene Wigner）

魏茨泽克（Carl Friedrich von Weizsäcker）

魏德涅夫斯基（Sebastian Wedeniwski）

魏兰德（Paul Weyland）

温伯格（Steven Weinberg）

温特伯格（Friedwardt Winterberg）

沃恩茨（Daniela Wuensch）

沃克（Evan Harris Walker）

乌伦贝克（George Uhlenbeck）

吴大猷

吴健雄

西格尔（Carl Ludwig Siegel）

希尔伯特 (David Hilbert)

肖恩哈格 (Arnold Schönhage)

谢希德

许良英

薛定谔 (Erwin Schrödinger)

雅默 (Max Jammer)

亚伯拉罕 (Max Abraham)

杨振宁

姚琦

英菲尔德 (Leopold Infeld)

约尔当 (Pascual Jordan)

约飞 (Abram Joffe)

詹姆士 (William James)

竺可桢

术语索引

BKS

EPR

X射线

β 衰变

阿哈罗诺夫-玻姆效应

阿塞斯腾教堂公墓

爱丁顿-芬克尔斯坦坐标

巴纳德奖章

贝尔不等式

毕安基恒等式

辩证法

波动力学

波尔约奖

波函数

波粒二象性

玻尔文献馆

不确定原理

超心理学

重整化

递归函数

递归可枚举集

电磁场方程

电磁观

电磁势

定态

丢番图方程

丢番图分析

度规

对应原理

非线性旋量理论

非因果联系原理

非因果平行性

菲尔茨奖

费马猜想

辐射

哥本哈根诠释

哥德巴赫猜想

哥德尔不完全性定理

格林尼治天文台

公理化

光电效应

光量子

光谱

广义相对论

广义相对论场方程

广义协变

规范变换

规范理论

黑体

互补性

互补原理

霍金-彭罗斯奇点定理

奇点定理

积分方程

极小值原理

几率

矩阵力学

康普顿效应

科摩演讲

科学革命综合症

拉格朗日四平方定理

兰姆达运算

勒波维茨不等式

黎曼 ζ 函数

黎曼-西格尔公式

黎曼猜想

黎曼张量

里奇张量

量子力学

量子场论

量子计算

量子纠缠

量子理论

量子密码

量子数

量子信息

临界线

零点

零点计算

逻辑

马列主义

米理论

纳粹主义

能级

能量动量守恒

欧拉四次方假设

判定问题

泡利效应

丘奇论题

散射

实在性

数论

斯塔克-爱因斯坦定律

四费米子相互作用

素数

素数定理

算法

索尔维会议

同位旋

沃尔夫斯凯尔奖金

希尔伯特-波利亚猜想

希尔伯特第十问题

狭义相对论

相对论

选择定则

杨-米尔斯理论

阴谋论

引力场方程

引力理论

有界全称量词

原子模型

跃迁

真空能

证伪

指数丢番图集

中微子

自旋

最小约束原理

最小作用量原理

作用量



原点阅读（The Origin），清华大学出版社旗下的图书品牌，秉承“科学，让个人更智慧，让社会更理性”的理念，致力于科学普及和科技文化类图书的出版，传播科学知识、科学精神、科学方法，展现科学的真实、独立、智慧、多变、宽容、动人及迷人。

理解科学丛书 卢昌海科普著作

FROM
SINGULARITIES
TO WORMHOLES

从奇点到虫洞

广义相对论专题选讲

卢昌海◎著

奇点、黑洞、白洞、虫洞、时间旅行……

对这些科普和科幻作品中迷人概念的深度探索；

霍金、彭罗斯、威顿、丘成桐、索恩……

对这些著名科学家精彩工作的细致解读。

清华大学出版社

目 录

[作者简介](#)

[第1章 能量条件](#)

[1.1 引言](#)

[1.2 能量条件](#)

[第2章 奇点与奇点定理](#)

[2.1 什么是奇点？](#)

[2.2 雷查德利方程](#)

[2.3 测地线束与共轭点](#)

[2.4 时空的因果结构](#)

[2.5 霍金-彭罗斯奇点定理](#)

[2.6 讨论](#)

[2.7 附录：雷查德利小传](#)

[第3章 正质量定理](#)

[3.1 渐近平直时空](#)

[3.2 广义相对论的动力学](#)

[3.3 ADM能量动量](#)

[3.4 正质量定理](#)

[3.5 舍恩与丘成桐的证明概述](#)

[3.6 威顿的证明概述](#)

[3.7 讨论](#)

[第4章 宇宙监督假设](#)

[4.1 黑洞、裸奇点及宇宙监督假设](#)

[4.2 摧毁黑洞——不可能的任务？](#)

[4.3 彭罗斯猜想与宇宙监督假设](#)

[4.4 壳层穿越奇点与壳层会聚奇点](#)

[4.5 走向严密表述](#)

[4.6 零质量标量场与裸奇点](#)

[4.7 讨论](#)

[第5章 虫洞物理学](#)

[5.1 萨根的小说](#)

[5.2 黑洞、白洞和虫洞](#)

[5.3 球对称可穿越虫洞](#)

[5.4 奇异物质——负能量的挑战](#)

[5.5 虫洞的“工程学”](#)

[5.6 由虫洞到时间机器](#)

[5.7 讨论](#)

[名词索引](#)

[人名索引](#)

[参考文献](#)

[后记](#)

[返回总目录](#)

作者简介

卢昌海，出生于杭州，本科就读于复旦大学物理系，毕业后赴美留学，于2000年获美国哥伦比亚大学物理学博士学位，目前旅居纽约。著有《那颗星星不在星图上：寻找太阳系的疆界》、《太阳的故事》和《黎曼猜想漫谈》，并曾在《中国青年报》、《数学文化》、《科幻世界》、《现代物理知识》、《中学生天地》、《科学画报》等报纸、杂志上发表几十篇科普及专业科普作品。

FROM
SINGULARITIES
TO WORMHOLES

从奇点到虫洞 ——广义相对论专题选讲

卢昌海 著

清华大学出版社
北 京

内容简介

本书以能量条件为线索，介绍普通教材中较少涉及或较少深入介绍的若干广义相对论专题，包括奇点及奇点定理、正质量定理、宇宙监督假设与虫洞物理学等。本书的介绍既包含背景或历史的介绍，也不回避必要的数学细节，融语言的生动风趣与内容的严谨翔实于一体，可使读者获得对这些专题的较深入的了解。

本书的介绍可作为普通教材的补充，供大学生、研究生及其他具有一定基础的物理爱好者阅读。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

从奇点到虫洞：广义相对论专题选讲 / 卢昌海著. —北京：清华大学出版社，2013

（理解科学丛书）

ISBN 978-7-302-32739-4

I. ①从... II. ①卢... III. ①广义相对论 IV. ①O412.1

中国版本图书馆CIP数据核字（2013）第130823号

责任编辑：邹开颜 赵从棉

封面设计：蔡小波

责任校对：赵丽敏

责任印制：杨 艳

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社总机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市金元印装有限公司

经 销：全国新华书店

开 本：165mm×240mm

印 张：9.75

字 数：136千字

版 次：2013年12月第1版

印 次：2013年12月第1次印刷

产品编号：048341-01

第1章

能量条件

1.1 引言

大家知道，广义相对论的场方程（即爱因斯坦场方程）[\(1\)](#)

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} \quad (1.1.1)$$

是一组有关时空度规的二阶非线性偏微分方程，求解这样的方程组是极其困难的。在20世纪60年代初以前，物理学家们对爱因斯坦场方程的很大一类研究都局限于在各种各样的简化条件——比如特定的对称性——下求解场方程。在这方面最著名的成果之一是德国物理学家施瓦西（Karl Schwarzschild, 1873—1916）于1916年得到的施瓦西解，其度规为

$$ds^2 = \left(1 - \frac{2m}{r}\right) dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2 d\Omega^2 \quad (1.1.2)$$

其中m为质量参数。另一个同样著名的成果则是俄国物理学家弗里德曼（Alexander Friedmann, 1888—1925）于1922年得到的弗里德曼解，其度规被称为罗伯逊-沃尔克度规（Robertson-Walker metric），为[\(2\)](#)

$$ds^2 = dt^2 - R^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right) \quad (1.1.3)$$

其中R为标度因子，k取值为0、-1或1，分别对应于平直（flat）、负常曲率（constant negative curvature）及正常曲率（constant positive curvature）空间。这两个度规分别是广义相对论在天体物理学和宇宙学上应用最为广泛的度规。

但这两个解的发现也带来了一个共同的问题，那就是它们所对应的

度规均具有奇异性。施瓦西度规是一个静态度规，它的奇异性（由上述表达式可以很容易地看到）出现在 $r=0$ 及 $r=2m$ 处。这其中 $r=2m$ 处的奇异性（一度被称为施瓦西奇点）后来被证明只是坐标选择导致的表观奇异性，可以通过坐标变换予以消除⁽³⁾；而 $r=0$ 处的奇异性则是真正的物理奇点，时空曲率在趋近该点时趋于发散⁽⁴⁾。这个奇点被称为曲率奇点。罗伯逊-沃尔克度规由于是动态度规，情形稍微复杂些。当 $k=1$ （即空间具有正曲率）时这一度规在 $r=1$ 处似乎具有奇异性，但这也是坐标选择导致的表观奇异性（读者们不妨自己寻找一个坐标变换来消去这一表观奇异性）。除去这一表观奇异性，从形式上看罗伯逊-沃尔克度规似乎没有其他显而易见的奇异性。但把这一度规代入到场方程中，研究它的动力学演化就会发现，对于我们观测到的膨胀宇宙来说，只要宇宙当前的物质分布满足一个很宽泛的条件，罗伯逊-沃尔克度规中的标度因子 $R(t)$ 在过去某个有限时刻就必定等于零。在那个时刻（通常定义为 $t=0$ ）宇宙的空间线度为零，物质密度则发散，因此那也是一个真正的物理奇点，被称为宇宙学奇点，或大爆炸（The Big Bang）。

这些奇点的出现是物理学家们所不乐意见到的，因为物理世界中并不存在真正意义上的无穷大。对于一个物理理论来说，出现无穷大往往意味着它的失效。因此奇点的出现对广义相对论是一种危机。不过当时物理学家们所知道的爱因斯坦场方程的解的数量十分有限，而且那些解大都具有很高的对称性（因为只有那种情形下的场方程才容易求解），比如施瓦西解具有球对称性，弗里德曼解则是均匀及各向同性的。这就给物理学家们提出了一个问题：由这几个特殊解所展示的危机究竟有多大的普遍性？或者说奇点的出现会不会只是那几个特殊解所具有的特殊对称性导致的特殊效应？如果是的话，那情势就不算太严重，因为那些对称性在现实世界里是不可能绝对严格地实现的，从而危机也就不具有普遍性。20世纪60年代，物理学家们对这一问题有两种不同的看法：一

种看法认为奇点的出现确实只是特殊对称性导致的特殊效应，如果考虑一般（即没有严格对称性）的情形，奇点将不会出现。持这种观点的代表人物是苏联物理学家栗弗席兹（Evgeny Lifshitz, 1915—1985）、卡拉特尼科夫（Isaak Markovich Khalatnikov, 1919— ）和贝林斯基（Vladimir Belinski, 1941— ）等。与之相反的另一看法则认为奇点在广义相对论中的出现是有普遍性的，从而并不是特殊对称性导致的特殊效应。持这种观点的代表人物是英国物理学家彭罗斯（Roger Penrose, 1931— ）和霍金（Stephen Hawking, 1942— ）等⁽⁵⁾。

这两组物理学家在奇点问题上不仅观点迥异，而且在这一领域所采用的研究方法也很不相同。栗弗席兹等人由于相信奇点的出现跟具体的解——尤其是其中的对称性——有关，因此把主要精力放在了求解一般——即没有严格对称性——情形下的场方程，以便探讨并检验在那种情形下是否不存在奇点；而彭罗斯和霍金等人也许是由于不认为奇点的出现跟具体的解有关，因此并不着眼于求解场方程，而大量运用了微分几何手段，通过所谓的“全局方法”（global method），在不直接求解场方程的情况下对奇点及奇点产生的条件进行了系统分析。如果说栗弗席兹等人的方法是正面强攻，那么彭罗斯和霍金等人的方法则属于旁敲侧击。经过几年的努力，两种方法分出了高下。栗弗席兹等人的正面强攻收效不大——因为爱因斯坦场方程实在太复杂了。虽然栗弗席兹等人的胃口并不贪婪，他们只研究宇宙学奇点 $t=0$ 附近的解而非全局性解，同时也并不奢望精确求解而采用了近似手段，但在不具有对称性的情形下，他们的努力依然遭到了难以逾越的困难⁽⁶⁾。另一方面，彭罗斯和霍金等人的“旁敲侧击”却获得了极大的成功，他们证明了一系列被称为奇点定理（singularity theorem）的著名结果，成为了经典广义相对论中登峰造极的成果之一。

不过彭罗斯和霍金等人的方法虽然不需要直接求解场方程，却也并非“不食人间烟火”，因为它与物质能量动量张量的性质依然有着密切关

系。这一点从物理上讲是显而易见的，因为正是物质能量动量张量的分布决定了时空的结构。爱因斯坦曾经把他的场方程比喻为一座建筑，这座建筑的一半是用精美的大理石（fine marble）砌成的，另一半却是用劣质的木料（low-grade wood）建造的。用精美的大理石砌成的那一半是方程的左端： $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ ，那是一个描述时空结构的优美的几何量，被称为爱因斯坦张量。而用劣质的木料建造的那一半则是方程的右端，也就是描述物质分布的能量动量张量： $8\pi T_{\mu\nu}$ 。为什么说这部分是用劣质木料建造的呢？因为自然界的物质分布种类繁多，物态方程千差万别，找不到一个普适的能量动量张量来描述所有已知的物质分布。不仅如此，在广义相对论所涉及的许多极端条件——比如某些星体内部的超高温、超高压、超高密度，宇宙演化的早期，以及引力坍缩的后期等条件——下还可能存在大量迄今未知的物质形态及分布。而且所有这些物质分布还可能在空间及时间上相互混合及转变。由于存在如此高度的复杂性，与爱因斯坦张量所具有的完全确定的数学结构相比，我们有关能量动量张量的知识无疑是极其贫乏的。

那么，在这种贫乏的知识下，如何才能研究诸如奇点的产生条件那样与物质的形态及分布密切相关，同时又具有很大普遍性的课题呢？彭罗斯和霍金等人采用了一种很高明的手法，那就是虽然谁也无法写下一个具有普适性的能量动量张量，但这一张量应当具备的某些基本条件（比如说能量密度必须大于等于零）在当时看来是具有很大的普适性的，因此他们假定物质的能量动量张量满足那些基本条件。另一方面，他们所使用的全局方法的威力之一就在于，只要利用那些基本条件，无须知道能量动量张量的具体形式，就可以得到许多非常有价值的结果。那些结果便是他们所证明的一系列奇点定理。而那些附加在能量动量张量上的条件则被统称为能量条件（energy condition）。如果说能量动量张量是用劣质木料建造的，那么能量条件的引进就好比是对那些劣质木

料套上几道铁箍进行加固，使它比原先的松散形式来得结实耐用。

1.2 能量条件

在本节中，将对几种主要的能量条件作一个简单介绍。不过，在介绍能量条件之前，我们首先要对能量动量张量本身的形式作以简单分析。在广义相对论研究中，为了让度规张量的形式尽可能简化，人们常常引进所谓的正交标架场（tetrad），对能量动量张量的分析也是如此。正交标架场（以下简称标架场）由一组正交归一的基矢量场 $(e_a)^\mu$ 组成，其中拉丁字母 a, b, \dots 标识标架场的基矢量，希腊字母 μ, ν, \dots 表示基矢量的时空指标⁽⁷⁾。标架场的基矢量满足下列正交归一条件：

$$\eta^{ab}(e_a)^\mu(e_b)^\nu = g^{\mu\nu}, \quad g_{\mu\nu}(e_a)^\mu(e_b)^\nu = \eta_{ab} \quad (1.2.1)$$

很明显，标架场不是唯一的，对一个标架场作局域洛伦兹变换所得到的仍然是标架场。由于洛伦兹群具有旋量表示⁽⁸⁾，因此标架场在讨论引力场与旋量场的相互作用时是非常重要的工具。对于我们所要讨论的能量条件来说，标架场的优点在于能量动量张量在标架场中的分量具有明确的测量意义。

霍金曾经把标架场下的能量动量张量分为四种类型，每种类型均可通过标架场中的洛伦兹变换约化为一个正则形式（canonical form）。这其中最重要的是第 I 类，其正则形式为

$$T^{ab} = \text{diag}(\rho, p_1, p_2, p_3) \quad (1.2.2)$$

其中 diag 表示对角矩阵， ρ 为标架场中的静止观测者（即世界线的切线沿基矢 e_0 方向的观测者）测量到的能量密度， p_i ($i=1, 2, 3$) 则为沿三个正交空间方向的主压强。除极少数特殊情形外，这种类型的能量动量张量涵盖了几乎所有物理上有意义的物质分布，我们在本书中将只讨论这种类型。

熟悉线性代数的读者可能会提出这样一个问题：第 I 类能量动量张

量的正则形式其实就是该张量的对角化，但能量动量张量是一个实对称张量，按照线性代数中熟知的定理，实对称张量必定可以通过正交变换对角化，既然如此，能量动量张量岂不都应该是第 I 类的？为什么在霍金的分类中它只是四种类型之一呢？这其中的原因在于普通线性代数所讨论的内积空间具有正定度规（positive definite metric），而广义相对论中的时空度规不是正定的（请读者想一想，度规的非正定性是如何破坏线性代数中有关实对称张量对角化的定理的？）。

下面我们就对几种主要的能量条件进行简单介绍：

弱能量条件（weak energy condition）：对所有类时矢量 V_a ， $T^{ab}V_aV_b \geq 0$ 。

利用 T^{ab} 的正则形式，我们可以证明：弱能量条件等价于 $\rho \geq 0$ 及 $\rho + p_i \geq 0$ （ $i=1, 2, 3$ ）。充分性的证明非常简单：取 $V_a = e_0$ （即静止观测者）可得 $\rho \geq 0$ ；取 $V_a \rightarrow e_0 + e_i$ （注意 V_a 是趋于而非等于 $e_0 + e_i$ ，因为后者是类光的）则可得 $\rho + p_i \geq 0$ 。接下来再证明必要性：假设 $\rho \geq 0$ 及 $\rho + p_i \geq 0$ ，则

$$T^{ab}V_aV_b = \rho V_0^2 + \sum_i p_i V_i^2 \geq \rho \left(V_0^2 - \sum_i V_i^2 \right) \geq 0 \quad (1.2.3)$$

其中第一个“ \geq ”用到了 $\rho + p_i \geq 0$ ，第二个“ \geq ”用到了 $\rho \geq 0$ 及 V_a 类时。

在弱能量条件中最重要的部分是 $\rho \geq 0$ ，它表明能量密度处处为正。需要提醒读者注意的是，虽然上述推导是在使正则形式成立的特殊标架场中进行的，但 $\rho \geq 0$ 这一结果适用于沿任意类时世界线运动的观测者所测得的能量密度（请读者想一想，这是为什么？）。由于物理上可以实现的所有观测者都是沿类时世界线运动的，因此弱能量条件表明任何物理观测者测得的能量密度都处处为正。

在弱能量条件中让 V_a 趋于类光，由能量条件的连续性可以得到：

零能量条件（null energy condition）：对所有类光矢量 k_a ， $T^{ab}k_ak_b \geq 0$ 。

显然（请读者自行证明），零能量条件等价于 $\rho + p_i \geq 0$ （ $i=1, 2, 3$ ）。零能量条件是一个非常弱的能量条件，比弱能量条件更弱。

强能量条件（strong energy condition）：对所有类时矢量 V_a ， $(T^{ab} - \frac{1}{2}g^{ab}T)V_aV_b \geq 0$ 。

由于爱因斯坦场方程可以改写为 $R^{ab} = 8\pi G [T^{ab} - (1/2)g^{ab}T]$ （其中 $T = T^a_a$ 为能量动量张量的迹），因此强能量条件等价于一个几何条件 $R^{ab}V_aV_b \geq 0$ ⁽⁹⁾。从物理上讲，强能量条件等价于 $\rho + \sum_i p_i \geq 0$ 及 $\rho + p_i \geq 0$ （ $i=1, 2, 3$ ）。这一点的证明非常简单，只需注意到在正则形式下：

$$T^{ab} - \frac{1}{2}g^{ab}T = \frac{1}{2} \begin{pmatrix} \rho + \sum_i p_i & & & \\ & \rho + 2p_1 - \sum_i p_i & & \\ & & \rho + 2p_2 - \sum_i p_i & \\ & & & \rho + 2p_3 - \sum_i p_i \end{pmatrix} \quad (1.2.4)$$

然后效仿前面有关弱能量条件的证明即可（请读者自行推导上式并完成论证）。

显然，强能量条件比零能量条件强。但是与强、弱二字的正常含义不符的是，强能量条件与弱能量条件互不包含，而非前者强于后者。事

实上，多数物质的主压强 p_1 是正的，对于那些物质，强能量条件其实比弱能量条件还弱⁽¹⁰⁾。

主能量条件 (dominant energy condition)：对所有类时矢量 V_a ， $T^{ab}V_aV_b \geq 0$ ，并且 $T^{ab}V_b$ 非类空。

这个能量条件是在弱能量条件之上增添了能流密度矢量 $T^{ab}V_b$ 非类空这一额外限制。在正则形式下这一额外限制可以表述为： $\|T^{ab}V_b\|^2 = \rho^2 V_0^2 - \sum_i p_i^2 V_i^2 \geq 0$ 。取 $V_b \rightarrow e_0 + e_i$ 可得 $\rho^2 \geq p_i^2$ 。这比弱能量条件中的 $\rho + p_i \geq 0$ 要强。为了证明 $\rho^2 \geq p_i^2$ 也是保证额外限制成立的充分条件，只需注意到

$$\|T^{ab}V_b\|^2 = \rho^2 V_0^2 - \sum_i p_i^2 V_i^2 \geq \rho^2 \left(V_0^2 - \sum_i V_i^2 \right) \geq 0 \quad (1.2.5)$$

这里第一个“ \geq ”用到了 $\rho^2 \geq p_i^2$ ，第二个“ \geq ”用到了 $\rho \geq 0$ 及 V_b 类时。将这一结果附加到弱能量条件上可得：主能量条件等价于 $\rho \geq |p_i|$ ($i=1, 2, 3$)。从定义及上述结果中均可看出，主能量条件比弱能量条件强（从而也比零能量条件强）。但它与强能量条件互不包含。

看到这里，有些读者可能会产生这样一个疑问：主能量条件中的额外限制是说能流密度矢量非类空。我们知道，在相对论中如果一个四维矢量类空，就必定可以找到一个参照系，使该矢量的时间分量为负。对于能流密度矢量来说，时间分量就是能量密度，因此如果能流密度矢量类空，就说明必定存在一个参照系，在其中能量密度为负。但弱能量条件已经表明任何物理观测者测得的能量密度都处处为正，这岂不等于排除了能流密度矢量类空的可能性？如果这样的话，主能量条件中的额外限制变成了弱能量条件的推论，这两种能量条件岂不就变成等价的了？这种推理显然是错误的，但它究竟错在哪里呢？有兴趣的读者不妨思考

一下，以加深对能量条件及其观测意义的理解。

迹能量条件 (trace energy condition) : $T \equiv T^a_a \geq 0$ 。

这是我们所要介绍的最后一种能量条件。它的表述与度规张量的符号约定有关，在本书中我们所用的约定是 $\eta_{ab} = \text{diag}(1, -1, -1, -1)$ 。如果做相反的约定，则迹能量条件的表述需改为 $T \leq 0$ 。在正则形式下，迹能量条件等价于 $\rho - \sum_i p_i \geq 0$ ，它与其他能量条件互不包含。

上面介绍的这几种能量条件有一个共同特点，那就是它们给出的都是每个时空点上能量动量张量所满足的条件，这样的能量条件被称为逐点能量条件 (pointwise energy condition)。除逐点能量条件外，人们还常常使用另外一类能量条件，称为平均能量条件 (average energy condition)，它们给出的是能量动量张量沿特定的时空曲线 (通常是类时或类光曲线) 所满足的平均意义上的条件。平均能量条件比相应的逐点能量条件弱，因为它们允许逐点能量条件在局部意义上被破坏，只要这种破坏能被所涉及的时空曲线上其他区段的贡献所弥补即可。

在本书接下来的各专题中，我们将以能量条件为线索，介绍广义相对论中一些重要、优美或有趣的课题，比如奇点定理、正质量定理 (positive mass theorem)、宇宙监督假设 (cosmic censorship hypothesis)、虫洞物理学 (wormhole physics) 等。我们将会看到，能量条件在所有那些课题中都有着重要应用^[11]。

注释

^[1] 在本系列的多数章节中，我们都取 $c=G=1$ 的单位制，并且不考虑宇宙学项。

^[2] 在弗里德曼之后，比利时天文学家勒梅特 (Georges Lemaître, 1894—1966)、美国物理学家罗伯逊 (Howard Robertson, 1903—1961)、英国数学家沃尔克 (Arthur Walker, 1909

—2001）均对这个解做过研究，因此这个解所对应的度规在文献中有若干不同的名称，最常见的是罗伯逊-沃尔克（RW）度规，也有称为弗里德曼-罗伯逊-沃尔克（FRW）度规或弗里德曼-勒梅特-罗伯逊-沃尔克（FLRW）度规的。

[\[3\]](#) 这一点最早是由勒梅特于1933年指出的。在那之前的1924年，英国天文学家爱丁顿（Arthur Eddington, 1882—1944）曾找到过一个在 $r=2m$ 处非奇异的坐标，但他并未意识到自己结果的重要性。消除了所有表观奇异性的施瓦西度规的最大解析延拓则是由美国数学家克鲁斯卡（Martin Kruskal, 1925—2006）和匈牙利数学家塞凯赖什（George Szekeres, 1911—2005）于1960年各自独立地给出的。

[\[4\]](#) 确切地讲，是由曲率张量构造出的某些标量——比如克莱茨曼标量（Kretschmann scalar） $R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}$ ——发散。这是与坐标选择无关的发散。

[\[5\]](#) 对这两种观点的区分（以及将彭罗斯和霍金视为后一种观点的代表人物）散见于文献中。但彭罗斯和霍金是从一开始（即在得到理论证据之前）就认为奇点的出现具有普遍性，还是在彭罗斯证明了第一个奇点定理之后才这样认为？笔者猜测是前者，但在文献中未见到说明。

[\[6\]](#) 栗弗席兹等人一度以为自己已对一般情况下奇点的不存在性作出了论述，并将结论写入了朗道（Lev Landau, 1908—1968）与栗弗席兹的名著《经典场论》（The Classical Theory of Fields）。20世纪60年代末，随着奇点定理的出现，他们意识到了自己的错误，并托到访苏联的美国物理学家索恩（Kip Thorne, 1940—）以最快的速度将更正错误的文章秘密带到了西方。不过，栗弗席兹等人虽未能达到他们希望达到的目标，但他们的工作为后人研究奇点附近的时空性质提供了重要参考，依然功不可没。

[\[7\]](#) 标架基矢（ e_a ） $^\mu$ 是时空坐标的函数，因此叫做标架场。“tetrad”这个名称通常是指四维的标架场（tetra这个词头的含意是“四”）。标架场的另一个常见的名称是vierbein，源于表示“四”的德语词头vier。在其他维数下，标架场还有其他一些常用名称，比如triad、pentad、funfbein、elfbein、vielbein等。

[\[8\]](#) 切空间中的一般线性变换群 $GL(4, \mathbb{R})$ 则没有旋量表示。

[\[9\]](#) 这里不考虑宇宙学项。另外，其他能量条件也可以用类似方式改写成几何条件。

[\[10\]](#) 有读者可能会问，既然强能量条件并不比弱能量条件强，那为什么会有这样的命名

呢？这是由于强能量条件可以写成 $T^{ab}V_aV_b \geq (1/2)T$ ，而弱能量条件为 $T^{ab}V_aV_b \geq 0$ ，由于通常 $T \geq 0$ ，因此如果把这两个能量条件视为是对 $T^{ab}V_aV_b$ 的约束条件，则强能量条件比弱能量条件强。当然，这种命名理由也是不严格的，因为 $T \geq 0$ 本身就是一种能量条件（即迹能量条件），而非无条件成立的物理事实。

[\[11\]](#)事实上，从历史的角度讲，能量条件的提出本身正是由那些课题促成的，因为正是在研究那些课题的过程中，物理学家们意识到需要对能量动量张量附加一定的约束，这种约束就是能量条件。

第2章 奇点与奇点定理

2.1 什么是奇点？

自本节开始，我们将介绍能量条件在广义相对论中的若干应用，首先要介绍的是奇点定理。在广义相对论中，对奇点的研究是一个重要的课题，它既是能量条件最早的应用之一，也是全局方法在广义相对论中初试锋芒的范例。我们在1.1节中曾经提到，广义相对论的经典解——比如施瓦西解——存在奇异性。这其中有的奇异性——比如施瓦西解中的 $r=2m$ ——可以通过坐标变换予以消除，从而不代表物理上的奇点；而有的奇异性——比如施瓦西解中的 $r=0$ ——则是真正的物理奇点。很明显，在奇点研究中，真正的物理奇点才是我们感兴趣的对象。

那么究竟什么是广义相对论中真正的物理奇点（简称奇点）呢？

初看起来，这似乎是一个很简单的问题。奇点显然就是那些时空结构具有某种病态性质（pathological behavior）的时空点。但稍加推敲，就会发现这种说法存在许多问题。首先，“病态性质”是一个很含糊的概念，究竟什么样的性质是病态性质呢？显然需要予以精确化。其次，广义相对论与其他物理理论有一个很大的差异，那就是其他物理理论都预先假定了一个背景时空的存在，因此，那些理论如果出现奇点——比如电磁理论中点电荷所在处的场强奇点，我们可以明确地标识出奇点在背景时空中的位置^{[\(1\)](#)}。但广义相对论所描述的是时空本身的性质。因此在广义相对论中一旦出现奇点，往往意味着时空本身的性质无法定义。另一方面，物理时空被定义为带洛伦兹度规（Lorentzian metric）的四维流形^{[\(2\)](#)}，它在每一点上都具有良好的性质。因此，奇点的存在本身就是与物理时空的定义相冲突的，物理时空按定义就是没有奇点的，或者

换句话说，奇点并不存在于物理时空中⁽³⁾。

既然奇点并不存在于物理时空中，自然就谈不上哪一个时空点是奇点，从而也无法把奇点定义为时空结构具有病态性质的时空点了。但即便如此，像施瓦西解具有奇异性这样显而易见的事实仍然是无可否认的，因此关键还在于寻找一个合适的奇点定义。

为了寻找这样的定义，我们不妨想一想，为什么即便将施瓦西解中的 $r=0$ 那样的“麻烦制造者”排除在物理时空之外，我们仍然认为施瓦西解具有奇异性是显而易见的事实？答案很简单（否则就不叫显而易见了）：当一个试验粒子在施瓦西时空中沿径向落往中心（即 r 趋于0）时，它所接触到的时空曲率趋于发散。由于试验粒子的下落是沿非类空测地线进行的⁽⁴⁾，这启示我们这样来定义奇点：如果时空结构沿非类空测地线出现病态性质，则表明存在奇点。这一定义不涉及奇点的位置，从而不需要将奇点视为物理时空的一部分，也就避免了上面提到的与物理时空的定义之间的冲突。但是，这一定义还面临两个问题：一是“病态性质”这个含糊概念仍未得到澄清；二是在这一定义中，假如试验粒子沿非类空测地线需要经过无穷长的时间才会接触到时空结构的病态性质的话，奇点的存在就不具有观测意义。为了解决这两个问题，物理学家们提出了一个进一步的要求，即要求定义中涉及的非类空测地线具有有限“长度”，并且是不可延拓的（inextendible）⁽⁵⁾。这种具有有限的“长度”的不可延拓非类空测地线被称为不完备非类空测地线（incomplete non-spacelike geodesics）。

有了不完备非类空测地线这一概念，我们可以这样来定义奇点：如果存在不完备非类空测地线，则时空流形具有奇点。这就是多数广义相对论文献所采用的奇点定义。这种存在不完备非类空测地线的时空被称为非类空测地不完备时空，简称测地不完备时空（geodesically incomplete spacetime）。在一些文献中，按照不完备测地线的类型，还将测地不完备时空进一步细分为类时测地不完备时空与类光测地不完备

时空^[6]。这一奇点定义的合理性体现在：在一个测地不完备的时空流形中，试验粒子可以沿不完备的非类空测地线运动，并在有限时间内从时空流形中消失。这种试验粒子在有限时间内从时空流形中消失的行为——即测地不完备性——可以视为是对时空结构具有“病态性质”这一含糊要求的精确表述。这样我们就既解决了“病态性质”的精确化问题，又使奇点具有了观测意义（即试验粒子在有限时间内就可以遇到奇点）。在一些文献中，还对奇点存在于过去还是未来进行区分：如果所涉及的非类空测地线是未来（过去）不可延拓的，则相应的奇点被称为未来（过去）奇点。

对于上述定义，还有一点需要补充。细心的读者可能注意到了，我们在引进“非类空测地线具有有限‘长度’”这一要求时，对“长度”一词加了引号。这引号所表示的是对长度定义的推广。具体地说，类时测地线的长度通常是被定义为固有时间的，即

$$\tau = \int ds \quad (2.1.1)$$

但这一定义不适合描述类光测地线，因为后者对应的固有时间恒为零。因此，为了描述类光测地线，我们需要对长度定义进行推广，推广为所谓的广义仿射参数（generalized affine parameter）。对于一条时空曲线 $C(t)$ （ t 为任意参数），广义仿射参数定义为

$$\lambda = \int \left[\sum_a V^a(t) V^a(t) \right]^{1/2} dt \quad (2.1.2)$$

其中 $V^a(t)$ 为曲线在 $C(t)$ 处的切向量 $\partial/\partial t$ 沿该处某标架场 $e_a(t)$ 的分量（曲线上各点的标架场定义为由某一点的标架场平移而来），求和是欧式空间中的分量求和。显然，这样定义的广义仿射参数是恒为正的，它的数值则与标架场的选择有关。不过可以证明，广义仿射参数的有限与否与标架场的选择无关，从而对于我们表述奇点的定义来说已经足够

了（因为我们只关心“长度”的有限性）。另外值得一提的是，广义仿射参数对所有 C^1 类（即一次连续可微）的时空曲线都可以定义，而不限于测地线。不难证明，类时测地线的固有时是广义仿射参数的特例（请读者自行证明）。

作为一个例子，我们来看看施瓦西解中 $r=0$ 的奇点是否满足上面所说的奇点定义。为此我们来计算从施瓦西视界（ $r=2m$ ）出发，向内（即沿 r 减小方向）延伸的径向类时测地线的长度（即固有时间）。由施瓦西度规可知

$$ds^2 = - \left(\frac{2m}{r} - 1 \right) dt^2 + \left(\frac{2m}{r} - 1 \right)^{-1} dr^2 \quad (2.1.3)$$

因此（请读者补全计算细节）

$$\tau = \int ds < \int \left(\frac{2m}{r} - 1 \right)^{-1/2} dr < \pi m < \infty \quad (2.1.4)$$

由此可见这种测地线的长度是有限的。另一方面，沿这种测地线趋近 $r=0$ 时，克莱茨曼标量 $R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}$ 发散，因此这种测地线是不可延拓的。这表明施瓦西解中 $r=0$ 的奇点满足上面所说的奇点定义。从物理上讲，这个结果表明落入施瓦西视界的试验粒子会在有限固有时间内从物理时空中消失（形象地说是“落入奇点”）。

现在让我们再回到定义上来，奇点的定义要求时空流形具有测地不完备性。读者也许会问：测地线究竟由于什么原因而不完备？另外，虽说测地不完备性是对时空结构所具有的病态结构的精确描述，但这“精确”二字是以数学上无歧义为标准的。在物理上，我们仍然可以问这样一个问题：当试验粒子沿不完备的测地线运动时，究竟会遇到什么样的时空病态性质？或者简单地说，奇点究竟是什么样子的？对此，物理学家们曾经试图给出直观描述，可惜一直没能找到一种直观描述足以涵盖所有可能的测地不完备性。比如，人们曾经认为奇点的产生意味着某些

几何量（比如克莱茨曼标量）或物理量（比如物质密度）发散，如果是这样，那么沿不完备非类空测地线运动的试验粒子所遇到的将是趋于无穷的潮汐作用或其他发散的物理效应。施瓦西奇点及大爆炸奇点显然都具有这种性质。但细致的研究发现，并非所有奇点都是如此。一个最简单的反例是锥形时空：

$$ds^2 = dt^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2) \quad (2.1.5)$$

其中 $r>0$ ， $0<\varphi<a$ （ a 为小于 2π 的一个角度），并且 $\varphi=0$ 与 $\varphi=a$ 粘连在一起。这个时空是局部平坦的（曲率张量处处为零），并且显然没有任何发散性。但这一时空无法延拓到 $r=0$ （被称为锥形奇点），因而是测地不完备的（类时与类光都不完备）[\[7\]](#)。这个反例表明奇点不一定意味着发散性。

对奇点的另一种直观描述是：奇点是时空中被挖去的点（或点集）。比如施瓦西奇点与刚才提到的锥形奇点是被挖去的 $r=0$ ，大爆炸奇点则是被挖去的 $t=0$ 。但这种描述如果正确的话，那么通向奇点的所有测地线——无论类时还是类光——必定全都是不完备的。换句话说，如果奇点是时空中被挖去的点（或点集），那么它的存在将同时意味着类时测地不完备性与类光测地不完备性。我们上面举出的所有例子都具有这一特点。但细致的研究表明，这一描述同样不足以涵盖所有的奇点。1968年，美国物理学家杰罗奇（Robert Geroch, 1942— ）给出了一个共形于闵科夫斯基时空的时空 $(R^4, \Omega^2\eta_{ab})$ ，其中共形因子 Ω^2 具有球对称性，在区域 $r>1$ 恒为1，在 $r=0$ 上满足 $t\Omega^2 \rightarrow 0$ （ $t \rightarrow \infty$ ）。显然（请读者自行证明），对于这样的时空，类时测地线 $r=0$ 沿 $t \rightarrow \infty$ 具有不完备性，因此这个时空流形具有类时测地不完备性。另一方面，所有类光测地线都将穿越区域 $r \leq 1$ 而进入平直时空，因而都是测地完备的。由此可见这一时空具有类时测地不完备性，但不具有类光测地不完备性[\[8\]](#)。这个反例表明奇点并非都能理解为是从时空中被挖去的点（或

点集）。

通过这些例子，我们对奇点定义所包含的复杂性有了一些初步了解，它的表述虽然简单，却巧妙地包含了难以完整罗列的种种复杂的时空类型。但另一方面，这个定义虽然已经具有很大的涵盖性，却仍不足以包含所有的奇点类型。这一点也是由杰罗奇指出的，此人在奇点定理的研究中是可以与霍金及彭罗斯齐名的非同小可的人物。1968年，在提出上述反例的同一篇论文中，杰罗奇给出了另外一种时空，它是测地完备的，但却包含长度有限的不可延拓类时曲线（注意是类时曲线而非类时测地线），并且该曲线上的加速度有界。从物理上讲，这意味着在这种时空中，具有有限燃料的火箭所携带的试验粒子沿特定的类时曲线运动，可以在有限时间之内从时空流形中消失。显然，这与自由下落的试验粒子从时空流形中消失具有同样严重的病态性质（事实上这里我们还要多损失一枚火箭！）。因此如果我们认为测地不完备性意味着奇点，那么就必须承认杰罗奇的时空也具有奇点。这个反例表明，我们——以及多数其他文献——所采用的测地不完备性只是定义奇点的充分条件，而不是必要条件。也就是说，一个测地不完备的时空必定具有奇点；但反过来则不然，一个测地完备的时空未必就没有奇点。

物理学家们对奇点性质所做的研究还有许多，限于篇幅，这里不再做进一步介绍，不过在后文介绍宇宙监督假设时我们还会再涉及这一话题。在接下来的几节中，我们将介绍奇点定理及其证明。

2.2 雷查德利方程

在上一节中我们对广义相对论中的奇点作了定义。这样定义的奇点究竟会在什么条件下出现？它是否如某些物理学家猜测的那样来源于对称性？这些都是奇点定理所要回答的问题。

由于我们对奇点的定义是建立在测地不完备性之上的，因此为了研究奇点产生的条件，很自然的做法就是对测地线的性质进行研究。我们用 V 表示测地线的切矢量，对于类时测地线来说， V 满足两个条件：

$V^a V_a = 1$ （归一化条件）及 $V^a V^b_{;a} = 0$ （自平移条件，其中“ $;$ ”为协变导数）。我们效仿线性代数中引进投影算符的做法，引进一个辅助张量 $h_{ab} = g_{ab} - V_a V_b$ 。不难证明（请读者自行验证）， h^a_b 是与 V 相垂直的子空间上的投影算符，因此 h_{ab} 有时被称为时空度规 g_{ab} 的“空间部分”（请读者想一想，这里所说的“空间”是什么含义？）。

我们知道，时空曲率的存在会导致沿相邻测地线运动的试验粒子之间的距离发生变化，这是所谓的测地偏离（geodesic deviation）效应，它是引力相互作用的一种体现。我们对测地线性质的研究也从这个角度入手，考察一个测地线束中的测地偏离效应。对一个测地线束来说，如果我们用与切矢量 V 相垂直的基矢 S 表示测地偏离矢量，则两者——作为矢量场——的对易子 $[S, V] = 0$ ，即（请读者自行证明）：

$dS^a/d\tau \equiv V^b S^a_{;b} = V^a_{;b} S^b$ （其中 τ 为固有时间）。这表明， $V^a_{;b}$ 描述了测地偏离矢量沿测地线的变化。如果我们把沿测地线束运动的一群粒子看成一种类似于连续介质的东西，那么 $V^a_{;b}$ 描述的就是这一连续介质的形变。由于这种形变是纯“空间”的（请读者想一想这是什么含义，并且予以证明），因此我们可以仿照连续介质力学的做法，用前面定义的时空度规的“空间部分” h_{ab} 将这种形变分解为（请读者加以验证）

$$V_{a;b} = \frac{1}{3}\theta h_{ab} + \sigma_{ab} + \omega_{ab} \quad (2.2.1)$$

其中 θ 、 σ_{ab} 及 ω_{ab} 分别定义为

$$\begin{cases} \theta = V_{a;b} h^{ab} = V^a_{;a} \\ \sigma_{ab} = V_{(a;b)} - \frac{1}{3}\theta h_{ab} \\ \omega_{ab} = V_{[a;b]} \end{cases} \quad (2.2.2)$$

这里 $V_{(a;b)}$ 与 $V_{[a;b]}$ 分别为 $V_{a;b}$ 的对称与反对称部分。上面这三项均有明确的物理意义： θ 被称为膨胀标量（expansion scalar），是 $V_{a;b}$ 的迹，描述的是测地线束会聚或发散的趋势； σ_{ab} 被称为切变张量

（shear tensor），是 $V_{a;b}$ 的无迹对称部分，描述的是测地线束的空间截面在体积不变（由无迹条件所保证）的情况下产生形变的趋势； ω_{ab} 被称为涡旋张量（vorticity tensor），是 $V_{a;b}$ 的反对称部分，描述的是测地线束在空间截面形状不变的情况下相互缠绕的趋势⁽⁹⁾。这其中描述测地线束会聚或发散的膨胀标量 θ 对于奇点定理的讨论有着特别重要的意义，因此我们将着重对它进行研究。

为了研究 θ ，我们注意到从物理上讲，影响 θ 的因素是时空曲率（或者说物质分布——两者通过爱因斯坦场方程彼此联系）。因此我们从曲率张量的定义式 $V^a_{;bc} - V^a_{;cb} = R^a_{dbc} V^d$ 出发⁽¹⁰⁾。将这一表达式对指标 a 和 b 进行缩并，与 V^c 取内积，并利用 $V_{a;b}$ 的分解式及类时切向量 V 的性质，便可证明 θ 沿测地线的变化为

$$\frac{d\theta}{d\tau} \equiv V^a \theta_{;a} = -R_{ab} V^a V^b - \frac{1}{3}\theta^2 - \sigma_{ab} \sigma^{ab} + \omega_{ab} \omega^{ab} \quad (2.2.3)$$

其中 τ 为固有时间。这个方程被称为雷查德利方程（Raychaudhuri

equation) [\(11\)](#), 是印度物理学家雷查德利 (Amal Raychaudhuri, 1923—2005) 与俄国物理学家朗道 (Lev Landau, 1908—1968) 彼此独立地提出的。雷查德利方程的提出恰好是在爱因斯坦逝世的那一年 (1955 年), 它与能量条件的结合将成为证明奇点定理的重要环节。

2.3 测地线束与共轭点

在雷查德利方程中，如果所考虑的测地线束局部正比于某个梯度场，或者说垂直于某个超曲面，则称该线束是超曲面垂直

（hypersurface orthogonal）的。可以证明，对于这样的测地线束来说，涡旋张量 ω_{ab} 为零，从而雷查德利方程可以简化为

$$\frac{d\theta}{d\tau} = -R_{ab}V^aV^b - \frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab} \quad (2.3.1)$$

由于 $\sigma_{ab}\sigma^{ab}$ 总是非负的，因此从这个方程中我们可以得到

$$\frac{d\theta}{d\tau} \leq -R_{ab}V^aV^b - \frac{1}{3}\theta^2 \quad (2.3.2)$$

如果进一步假定强能量条件成立，即 $R_{ab}V^aV^b$ 处处非负，则上述不等式可以进一步简化为

$$\frac{d\theta}{d\tau} \leq -\frac{1}{3}\theta^2 \quad (2.3.3)$$

对这个不等式进行积分可得

$$\theta^{-1} \geq \theta_0^{-1} + \frac{1}{3}(\tau - \tau_0) \quad (2.3.4)$$

其中 $\theta_0 = \theta(\tau_0)$ 。从这个不等式我们可以得到一个重要的推论，那就是倘若 $\theta_0 < 0$ ，即线束在 $\tau = \tau_0$ 时出现会聚效应，则 θ 会在有限固有时间 $\tau - \tau_0 \leq 3/|\theta_0|$ 内趋于负无穷。可以证明，这意味着测地线束在该处会聚为一点，或者说测地偏离矢量场——也称为雅可比场（Jacobi field）——在该处为零。如果一个从 p 点发出的非平凡（即各测地线不处处重合，或者说雅可比场不处处为零）的类时测地线束在 q 点会聚，我们就把 q 和 p

称为该测地线束上（即其中每一条测地线上）的一对共轭点（conjugate points）。从上面的分析中我们看到，如果从 p 点发出的一个类时测地线束在未来某一点上出现会聚效应 $\theta < 0$ ，则在该线束上距离 p 有限远的地方必定存在一个与 p 共轭的点 q ——当然，这里我们要假定该测地线束可以延伸到 q 点。

显然，在一个测地完备时空中，“测地线束可以延伸到 q 点”这一假定是自动满足的。因此，对于测地完备时空来说，上面这个结果是所有类时测地线都满足的普遍性质。进一步的分析表明，上述结果所要求的条件，即测地线束在“某一点上出现会聚效应 $\theta < 0$ ”，可以转化为一个有关曲率张量的条件。事实上，由式（2.3.1）和式（2.3.4）可以看到，即便在 $\sigma_{ab}\sigma^{ab}$ 与 $R_{ab}V^aV^b$ 处处为零（此时式（2.3.4）取等号形式），且 $\theta_0 > 0$ 这一对于形成 $\theta < 0$ 来说最为不利的条件下， θ 仍将在 $t \rightarrow \infty$ 时趋于零（即几乎就要形成 $\theta < 0$ 这一结果）。这使人想到，上述最为不利的条件只要在某一点上（从而由连续性条件可知在该点的一个邻域内）被破坏，比如 $R_{ab}V^aV^b > 0$ 在某一点上成立，就足可造成当 τ 足够大时 $\theta < 0$ 。可以证明，事实的确如此。因此“某一点上出现会聚效应 $\theta < 0$ ”这一条件可以转化为某一点上 $R_{ab}V^aV^b > 0$ 。如果我们进一步把 $\sigma_{ab}\sigma^{ab}$ 所起的作用也考虑进去，这一条件还可以继续减弱，最终可以得到这样一个结果：在一个测地完备的时空中，如果强能量条件成立，并且在每条类时测地线上至少有一个点使得 $R_{abcd}V^bV^d \neq 0$ ，则所有类时测地线上都存在共轭点对，简称共轭对。

从物理意义上讲，每条类时测地线上至少有一个点使得 $R_{abcd}V^bV^d \neq 0$ ，意味着每条类时测地线都至少会在一个时空点上遇到由物质分布或引力波所造成的某种测地偏离效应。这一条件——称为类时一般性条件（timelike generic condition）——在理论上可以被一些非常特殊的情形（比如曲率张量与测地线切矢量形成特殊分量匹配的情形）所违反。但对于具有现实物理意义的情形来说，由于物质及引力波的分

布往往足够弥散及随机，类时一般性条件被认为是得到满足的。

上面这些结果都是针对类时测地线的。不过可以证明，除了一些不影响定性结果的差异（比如固有时间 τ 变成仿射参数 λ ，雷查德利方程中的数值因子 $1/3$ 因垂直子空间维数的改变而变成 $1/2$ 等）外，类光测地线也具有类似性质。类光测地线所满足的一般性条件为：每条类光测地线上至少有一个点使得 $k_{[e} R_{a]bc[d} k_{f]} k^b k^c \neq 0$ 。这个条件被称为类光一般性条件（null generic condition）[\(12\)](#)。类时与类光一般性条件统称为一般性条件[\(13\)](#)。把类时与类光情形合在一起，我们前面所介绍的结果可以重新表述为：在一个测地完备的时空中，如果强能量条件与一般性条件成立，则每条非类空测地线上都存在共轭对[\(14\)](#)。这是一个不依赖于对称性的普遍结果，它对于奇点定理的证明及确立奇点定理的普适性都有极其重要的作用。

细心的读者可能还记得，在上述结果的证明伊始，我们曾经作过一个假设，即所考虑的测地线束是超曲面垂直的。这个假定保证了 $\omega_{ab} = 0$ ，从而消除了雷查德利方程中与其他各项符号相反——因而会对我们的证明造成极大干扰——的 $\omega_{ab}\omega^{ab}$ 项（请读者想一下，这一项的符号与其他各项相反的物理意义是什么？）。那么这个假设具有多大的普遍性呢？或者说，这个假设是否会使上述结果——进而使整个奇点定理的证明——失去应有的普遍性呢？答案是否定的，因为在数学上可以证明，经过某一时空点的类时测地线束必定在该点的某个凸邻域内具有超曲面垂直性，因此 ω_{ab} 在该邻域内必定为零。不仅如此，通过一个与雷查德利方程类似的描述 ω_{ab} 沿测地线变化的方程可以证明， ω_{ab} 沿一条测地线只要在某一点上为零，就沿该测地线处处为零。因此，假定测地线束为超曲面垂直不会有损结果的普遍性。

综合上述分析，我们看到，在一个具有适当物质分布的测地完备时空中共轭点的存在是普遍现象。假如有一个适当的物质粒子群沿某个非

类空测地线束运动，那么当它们运动到共轭点上时，由于测地线的会聚，粒子的数密度（以及质量密度）将趋于发散，从而形成一个奇点。雷查德利发表于1955年的原始论文就涉及了这样的情形⁽¹⁵⁾。不过，在一般情况下显然并没有理由假定存在那样的物质粒子群，因此共轭点的存在只是一个抽象的几何结果，不会直接导致奇点，上述结果也不足以作为奇点存在性的证明（如果一定要算证明的话，只能算是非常弱的证明，因为它所要求的条件太过特殊）。但是，这一结果为十年后彭罗斯等人的工作奠定了基础，是证明奇点定理的第一步。这一步所侧重的是引力理论中的动力学因素，强能量条件的引进是这种因素的体现。

在下一节中将会看到，当我们把有关测地线的上述结果与看似风马牛不相及的时空的因果性质结合起来时，奇点在广义相对论中的出现就变得几乎不可避免了。

2.4 时空的因果结构

在前两节中，我们介绍了证明奇点定理的第一步，即通过雷查德利方程、强能量条件及一般性条件，确立了测地完备时空中每条非类空测地线上都存在共轭对这一结论。那一步侧重的是引力理论中的动力学因素。接下来的第二步侧重的则是时空的因果结构。

在物理学中，因果性是一个很微妙的概念。一方面，它是现实世界中最基本的经验事实之一；另一方面，却很少有物理理论直接把因果性作为前提条件。由此导致的一个结果是：某些物理理论起码在形式上允许因果性的破坏。广义相对论就是这样的理论。在广义相对论中，破坏因果性最简单的方式是产生闭合非类空曲线（请读者想一想，这种曲线在什么意义上破坏因果性？）。为了对这种类型的因果性破坏进行界定，人们引进了一个条件，叫做因果性条件（causality condition）：一个时空如果不存在闭合非类空曲线，则称为满足因果性条件。考虑到所有有质量粒子都只能沿类时曲线运动，人们还提出了一个比因果性条件稍弱的条件，称为时序条件（chronology condition），它与因果性条件的差别在于把“不存在闭合非类空曲线”减弱为“不存在闭合类时曲线”。

虽然时序条件比因果性条件稍弱，但可以证明，一个时空如果在时序条件之外还满足测地完备性，则该时空将不仅满足因果性条件（即不存在闭合非类空曲线），而且不存在可以无限逼近闭合非类空曲线的曲线。这个比因果性条件更强的性质被称为强因果性条件（strong causality condition），它在奇点定理的研究中是一个重要概念。

奇点定理研究中的另一个重要概念是所谓的封闭陷获面（closed trapped surface）。这是一种特殊的二维封闭类空曲面，其基本性质是：所有与之正交的类光测地线束无论向内还是向外都是趋于会聚的（即膨

胀标量 $\theta < 0$)。从物理上讲,该性质意味着从封闭陷获面发出的光波的波前是收缩的。这种曲面在广义相对论中并不鲜见,比如施瓦西解中所有 $r < 2m$ 的曲面都具有这一性质(这表明任何物质——包括光波——都不能从施瓦西黑洞中逃脱)。1983年,美国数学家舍恩(Richard Schoen, 1950—)与美籍华裔数学家丘成桐(Shing-Tung Yau, 1949—)证明了一个相当普遍的结果:只要物质的分布足够致密,就必定会出现封闭陷获面。

由于封闭陷获面的定义是建立在类光测地线的行为之上的,因此我们引进与类光测地线有关的两个特殊点集: $E^+(S)$ 与 $E^-(S)$,分别由从曲面 S 发出的未来与过去方向的类光测地线所组成^[16]。在这一定义中我们还进一步假定 S 上任意两点之间都不存在类时连接,这种点集 S 被称为非时序点集(achronal set)。封闭陷获面由于是类空的,显然也是非时序点集。可以证明,如果强能量条件成立,则对于任何封闭陷获面 S , $E^+(S)$ 与 $E^-(S)$ 紧致。

我们知道,物理上所有的相互作用都是非类空传播的(也就是说相互作用的传播速度不大于光速)。因此,如果我们考虑时空中某一点上的任何物理性质,它所能依赖的初始条件只能位于与该点具有非类空连接的时空点上。反过来说,给定某个时空区域 S 上的初始条件,我们能完全确定其性质的时空区域是由那样的一些点组成的:所有通过那些点的过去不可延拓非类空曲线都与 S 相交。这一时空区域被称为 S 的未来柯西展开(future Cauchy development)或未来影响域(future domain of dependence),通常记为 $D^+(S)$ 。 $D^+(S)$ 的边界则被称为未来柯西视界(future Cauchy horizon),记为 $H^+(S)$ 。类似地,我们也可以定义 S 的过去柯西展开(或过去影响域)和过去柯西视界,分别记为 $D^-(S)$ 和 $H^-(S)$ 。 S 的未来柯西展开与过去柯西展开合在一起——即 $D^+(S) \cup D^-(S)$ ——称为 S 的柯西展开(或影响域),记为 $D(S)$ 。一个时空(或时空中的一个点集) M 中如果存在一个封闭非时

序点集 S ，使得 $M=D(S)$ ，则称为是全局双曲（globally hyperbolic）的，相应的封闭非时序点集 S （可以证明它一定是一个超曲面）被称为柯西面（Cauchy surface）。柯西面可以被形象地理解为时空中对应于某一时刻的超曲面。一个时空如果是全局双曲的，我们就可以通过柯西面上的初始条件预言整个时空中的演化，因此时空的全局双曲是一种非常优良的因果性质。1965年，彭罗斯正是在假设时空为全局双曲的基础上证明了最早的奇点定理。但是，时空的全局双曲是一个很强的假设，要想证明现实时空满足这样的假设几乎是不可能的。因此五年之后（即1970年），霍金与彭罗斯放弃了这一假设，在一组物理上更容易实现的假设之上重新证明了奇点定理，这便是我们所要介绍的霍金-彭罗斯奇点定理。

对于奇点定理的证明来说，全局双曲时空（或点集）有一个很重要的性质，那就是其中任意两个可以建立非类空连接的时空点 p 和 q 之间必定存在一条非类空测地线，其长度最大——即大于或等于 p 和 q 之间的任何其他非类空曲线（请读者想一想，为什么这里测地线的长度是最大而不是最小？），并且在 p 和 q 之间不存在与 p 共轭的点。

2.5 霍金-彭罗斯奇点定理

以上我们介绍了一些在奇点定理的研究中常用的有关时空因果性质的定义及结果。虽然介绍得比较零散，但有些读者可能已经看出一点思路来了：我们在2.3节中曾经证明了，在适当的条件下，每条非类空测地线上都存在共轭对；而在上节的末尾我们则开始接近一个与之矛盾的结果，即在特定的条件下，某些测地线上不存在共轭对。这对彼此矛盾的结果正是证明奇点定理的关键。确切地讲，奇点定理的证明是要通过这对彼此矛盾的结果来论证以下5个条件不可能同时成立：

- (1) 时空是测地完备的；
- (2) 强能量条件成立；
- (3) 一般性条件成立；
- (4) 时空满足时序条件；
- (5) 时空中存在一个非时序点集 S ，使得 $E^+(S)$ 与 $E^-(S)$ 紧致。

限于篇幅，我们只能简单叙述一下论证思路。在上述5个条件中，条件(1)～(3)是2.3节所介绍的证明奇点定理的第一步中所用到的条件，由此推知的是每条非类空测地线上都存在共轭对。条件(1)和条件(4)所推知的——如上文所述——是时空满足强因果条件。而由强因果条件与条件(5)则可以证明这样一个结果：时空中存在一个全局双曲区域 M ，其中包含一条未来不可延拓类时曲线 γ 及一条过去不可延拓类时曲线 λ 。利用这一结果就可以证明时空中存在一条没有共轭对的非类空测地线。具体做法是：在 λ 上取一个沿过去方向趋于无穷的点集 a_n ，同时在 γ 上取一个沿未来方向趋于无穷的点集 b_n （选取时使得 b_1 在 a_1 的类时未来，从而保证所有 b_n 都在 a_n 的类时未来）。由于 M 是全局双曲的，因此——由上节末尾的结果可知——在每一对 a_n 和 b_n 之间都存在

一条（长度最大的）非类空测地线 μ_n ，其上在 a_n 和 b_n 之间不存在 a_n 的共轭点。可以证明，存在于 M 之中的这一由非类空测地线 μ_n 组成的无穷集合必定有一个聚点（limit point） μ ，它是一条非类空测地线，并且其上不存在任何共轭对。这样，我们就得到了与第一步（也就是条件（1）～（3））所得的“每条非类空测地线上都存在共轭对”相矛盾的结论，从而证明了上述5个条件不可能同时成立。

既然上述5个条件不可能同时成立，那么我们就可以选其中4个条件为前提（即假定这4个条件成立），来推翻剩下的那个条件⁽¹⁷⁾。霍金与彭罗斯所做的是以条件（2）～（5）为前提，来推翻条件（1），即证明时空不是测地完备的。按照我们在2.1节所作的定义，这表明时空中存在奇点。这就是霍金与彭罗斯的奇点定理。

在被奇点定理采用为前提的条件（2）～（5）中，条件（2）～（4）都有明确的物理意义，唯独条件（5）——即时空中存在一个非时序点集 S ，使得 $E^+(S)$ 与 $E^-(S)$ 紧致——显得很抽象。幸运的是，我们可以用一些物理意义更为明确的条件来取代这一抽象的数学条件。在上节中我们介绍过，如果强能量条件成立，则对于任何封闭陷获面 S ， $E^+(S)$ 与 $E^-(S)$ 紧致。由于强能量条件已经包含在条件（2）～（4）中了（即条件（2）），因此我们可以用“时空中存在封闭陷获面”来替代条件（5），这个条件在物理上可以由足够致密的星体来满足。除此之外，霍金与彭罗斯还提出了另外两个条件来替代条件（5）：一个是“时空中存在紧致无边的非时序点集”⁽¹⁸⁾，这个条件在物理上可以由空间上有限无边的宇宙来满足；另一个是“时空中存在一个点，通过该点的所有未来（或过去）方向的类光测地线束的膨胀标量 θ 最终将变为负值”，这个条件在物理上可以由局部膨胀或收缩的宇宙来满足。这3个替代条件都被认为是原则上可以检验，并且很可能在我们的宇宙中已经得到满足的条件。

至此，我们可以对霍金与彭罗斯所证明的奇点定理做一个完整表

述：

霍金-彭罗斯奇点定理：一个时空若满足以下条件，就必定是非类空测地不完备的（即存在奇点）：

（1）强能量条件成立。

（2）一般性条件成立。

（3）满足时序条件。

（4）以下三个条件之一成立：

①存在封闭陷获面；

②存在紧致无边非时序点集；

③存在一个点，通过该点的所有未来（或过去）方向的类光测地线束的膨胀标量 θ 最终将变为负值。

这个定理是霍金与彭罗斯于1970年提出并证明的。如我们在上文中所说，这并不是最早的奇点定理。彭罗斯于1965年、杰罗奇于1966年、霍金于1967年都提出过奇点定理。比较之下，霍金-彭罗斯奇点定理所要求的条件在物理上最容易实现，并且涵盖面也广^[19]，因此人们如今提到奇点定理时通常指的就是这一定理。霍金-彭罗斯奇点定理不依赖于对称性，它对于确立广义相对论中奇点的存在性及普遍性是非常重要的。同时它也是对我们1.1节中所介绍的有关奇点的不同见解的有力裁决。但是，霍金-彭罗斯奇点定理也有一个显而易见的缺点，那就是它既无法告诉我们究竟哪一条非类空测地线是不完备的，也无法提供有关奇点具体性质的信息。这一缺点为后人加强奇点定理的结论部分留下了空间。不过要想加强奇点定理的结论部分，往往不可避免地要对前提部分也予以加强，从而有损定理的普遍性。

2.6 讨论

我们对奇点定理的介绍就要结束了。有些读者可能会提出这样一个问题：我们证明霍金-彭罗斯奇点定理所用的是排除法，即通过证明测地完备性与奇点定理的4个前提不相容，来排除测地完备性，从而确立奇点的存在。但是，当一组命题不相容时，究竟哪个（或哪几个）命题应该被排除，在逻辑上是有很随意性的⁽²⁰⁾。因此从逻辑上讲，由上面介绍的不相容性，原则上可以通过排除不同的命题，而得到不同的定理（读者不妨自己写出几个看看）。为什么我们偏偏要选择将时空的测地完备性作为被排除的命题，从而得到霍金-彭罗斯奇点定理呢？

这是一个非常好的问题。我们知道，一个物理上有价值的定理必须能对物理世界作出某种程度的描述。因此，在所有逻辑上成立，并且能进行物理诠释的数学命题中，只有那些其前提在物理上能够实现的定理才能成为有效的物理定理。如果已经知道物理世界不满足某一性质，那么把该性质作为前提的数学命题就不能成为有效的物理定理。从这个意义上讲，我们可以通过考察霍金-彭罗斯奇点定理所涉及的4个前提在物理世界中实现的可能性，来分析这一定理的合理性。

在霍金-彭罗斯奇点定理的4个前提中，前提（4）属于初始及边界条件，并且实现的可能性极大。事实上，早在霍金-彭罗斯奇点定理提出的年代，天文观测及理论研究就已经在很大程度上显示出这个前提的3个子条件很可能部分甚至全部得到满足。前提（1）和前提（2）与人们在宏观世界的观测经验相符，因为迄今所知的所有宏观物质的能量动量张量都满足强能量条件，而现实宇宙中物质（包括宇宙微波背景辐射）及引力波的分布无疑遍及全空间，从而满足一般性条件，因此在以大尺度宏观世界为主要描述对象的广义相对论中，这两个前提被认为是适用的。前提（3）所要求的不存在闭合类时曲线也具有不错的经验基

础，因为时间的单向性是宏观世界上最基本的经验事实之一。因此所有这4个前提都有可信赖之处，这在很大程度上保证了霍金-彭罗斯奇点定理的价值。

但如果一定要在这4个前提中找出一个最有可能在现实物理世界中不成立的，那么——如我们将在后文中看到的——能量条件（即前提（1））将是首选，因为理论与观测都表明它事实上就不成立。不过，能量条件的破坏主要来自量子效应，而我们所讨论的奇点定理是经典广义相对论中的命题，两者在所涉范围上是有出入的。那么，假如我们不考虑量子效应，或者说只考虑经典广义相对论，又有哪一个前提最值得怀疑呢？一般认为是时序条件（即前提（3））。这一条件要求不存在闭合类时曲线。它之所以值得怀疑，主要有两个原因：一是因为广义相对论的某些特殊解事实上允许闭合类时曲线存在，虽然迄今为止那些解还没有一个得到过任何观测上的支持；二是由于闭合类时曲线实际上是一种抽象的时间机器，这是一种在很多方面都很引人入胜的东西。因此有些物理学家把广义相对论没有在原理层面上禁止闭合类时曲线，视为是一个很值得探索的理论问题。

如果时序条件有可能被破坏，那就产生了一个很自然的问题：我们是否可以通过作一个与霍金-彭罗斯奇点定理不同的选择，把测地完备性作为定理的前提之一，而把时序条件的破坏（从而允许时空中存在闭合类时曲线）作为定理的结论呢？[\(21\)](#)对这种可能性物理学家们也进行过一些研究。1977年，美国图兰大学（Tulane University）的物理学家梯普勒（Frank Tipler, 1947— ）研究了渐近平直时空中有限大小的闭合类时曲线，结果发现在强能量条件与一般性条件等成立的情况下，这样的曲线在测地完备时空中是不可能出现的[\(22\)](#)。其他一些物理学家后来也做了这方面的研究和推广，包括使用更弱的条件，以及推广时序破坏的定义等，得到的结果都类似。这些结果成为后来霍金提出所谓时序保护假设（chronology protection conjecture）的基础之一。这些结

果表明，时序条件的破坏在很大程度上本身就意味着测地完备性的破坏，因而放弃时序条件并不能挽回测地完备性⁽²³⁾。这在一定程度上进一步加强了奇点的不可避免性⁽²⁴⁾，也进一步支持了霍金-彭罗斯奇点定理的合理性——当然，所有这一切都限于经典广义相对论的范围。

2.7 附录：雷查德利小传

2005年6月18日，印度物理学家雷查德利（见图2.1）因心脏病发作猝然去世，享年81岁。雷查德利是为数不多的名字进入科学术语的印度物理学家之一，但他的名字对于多数人——包括他的印度同胞以及世界各地广义相对论专业以外的物理系学生——来说都显得比较陌生，他的生平更是鲜为人知。在这里，我们将对雷查德利的生平作一个简单介绍，作为奇点与奇点定理介绍的背景资料。同时笔者也希望本节能对以著名科学家为主要关注对象的中文传记资料起到一定的补充作用。从某种意义上讲，了解一些像雷查德利这样的“非著名科学家”的生平，对于今后从事科学研究的学生来说，或许会有一些从名人传记中无法得到的启示——因为他们的人生和际遇更接近常人。



图2.1 印度物理学家雷查德利（1923—2005）

雷查德利于1923年9月14日出生在巴里萨尔（Barisal），这是当时英属印度（British India）的一部分，如今则是孟加拉国中南部的一个货物转运中心。雷查德利的父亲是印度东北部大城市加尔各答（Calcutta）一所初级学校的数学教师，这座城市后来也成为雷查德利

一生主要的学习和工作地点。或许是受父亲的影响，雷查德利自小就对数学怀有兴趣，并且成绩优异。

中学毕业后，雷查德利进入了印度著名学府加尔各答大学（University of Calcutta）附属的院长学院（Presidency College）。他当时最感兴趣的是数学，其次则是物理。不过在选择专业这个节骨眼上，雷查德利的父亲再次对他产生了影响。在父亲的建议下，物理成为了雷查德利的专业。一位数学教师有一位对数学感兴趣的儿子，他为什么不让儿子“子承父业”呢？这是因为雷查德利的父亲虽只是一名不起眼的初级学校教师，却有着当时极为出众的数学硕士学位，他的这种大材小用的郁闷经历使得他希望自己的儿子能尝试一个不同的专业。

1942年，雷查德利顺利地获得了学士学位，两年后又获得了加尔各答大学的硕士学位，平了父亲的学位纪录。1945年，雷查德利成为印度科学培训协会（India Association for the Cultivation of Science, IACS）的一名研究人员，开始了自己的学术生涯。印度科学培训协会虽然有着很土的名字，却是印度历史最悠久的基础科学研究院。印度最著名的本土物理学家拉曼（Chandrasekhara Raman, 1888—1970）正是在这里发现了以他名字命名的拉曼效应（Raman effect），并获得1930年的诺贝尔物理学奖。但对雷查德利来说，进入声名卓著的科学培训协会并未使自己的学术之路成为坦途。那时候，他对广义相对论产生了兴趣，但科学培训协会却要求他从事实验工作，因为广义相对论在当时已被视为是很难做出新成果的学科。由于谋职不易，雷查德利痛苦地服从了分配，但他凭借自己的兴趣自学了微分几何与广义相对论。四年后，雷查德利在加尔各答大学附属的阿属托什学院（Asutosh College）获得了一个临时教职。在那期间，他对广义相对论中施瓦西解的奇点进行了研究，这一研究在很大程度上预示了他一生的兴趣^{[\[25\]](#)}。

雷查德利在广义相对论方面的研究没能使他在阿属托什学院的临时职位变为永久。1952年，他重新回到了印度科学培训协会。距他上一次

进入该协会，时间虽已相隔七年，但广义相对论不被看好的局面依然如故。雷查德利此次被要求从事的是金属性质研究。与在阿属托什学院的职位一样，雷查德利在科学培训协会的位置也是临时的，需要发表一定数量的论文来维持。不过，在这种被迫在一个自己不感兴趣的领域从事工作的压力干扰下，雷查德利并未放弃对广义相对论的研究。1955年，他在一篇研究奇点问题的文章中提出了如今以他名字命名的雷查德利方程⁽²⁶⁾。他的这一工作为十年后彭罗斯与霍金等人证明奇点定理做了重要铺垫，也为他所热爱的广义相对论重新成为研究热点起到了间接作用⁽²⁷⁾。不过，雷查德利的这一工作在印度并未引起反响，最早的支持来自于遥远的欧洲，据说量子力学创始人之一的约尔当（Pascual Jordan, 1902—1980）曾在一个讨论会上称赞雷查德利的工作。比雷查德利稍晚，著名俄国物理学家朗道也独立地提出了雷查德利方程。此后几年，雷查德利方程的重要性渐渐受到一些广义相对论学者的注意，这一成功给了雷查德利很大的信心和勇气，他以此为基础撰写了博士论文。雷查德利的博士论文受到了著名广义相对论学者惠勒（John Wheeler, 1911—2008）的好评。1959年，他获得了物理学博士学位⁽²⁸⁾。

1961年，雷查德利的母校院长学院为他提供了一个理论物理教授的职位，雷查德利回到了阔别二十年的熟悉校园，那里从此成为了他的学术归宿。在此后的漫长岁月里，雷查德利曾几度申请研究所的职位，但都没有成功。在学术生涯的主要时间里，他几乎没有获得任何荣誉，以至于一些印度物理学家在他去世后沉痛地将他称为最受忽视的印度物理学家。雷查德利在院长学院期间为印度的物理教育做出了卓越贡献。印度独立后，为促进科研，追赶西方国家在科技领域的领先地位而建立了不少纯研究性的机构。那些机构具有非常独立的地位，设备与条件均优于大学，因而吸引了大量的研究人员。很多年后，印度科学界开始反思那些机构对印度高等教育的负面影响。简单的说，那些机构对研究人员

的吸积作用，以及它们与高校完全脱钩的体制，使得印度第一流的研究人员脱离了学生，而第一流的学生又找不到好的教师。在那样的环境下，雷查德利默默无闻地“陷落”在高校，反倒成为了印度物理学界的幸事。雷查德利的一些学生后来成为了印度重要的物理学家，其中包括像森（Ashoke Sen，1956— ）这样具有国际知名度的理论物理学家。雷查德利去世后的第二天，他的一位学生在给同事的电子邮件中写道：“我们中有幸在本科阶段成为他学生的许多人，都把他视为我们一生遇到的最伟大的物理老师。”

1972年，印度天体物理学家纳里卡（Jayant Narlikar，1938— ）从英国剑桥“海归”到了印度的塔塔基础研究所（Tata Institute of Fundamental Research）。他的回国不仅将一些前沿的物理信息带回了印度，也把雷查德利方程在西方国家广义相对论研究中所起的作用反馈回了印度。雷查德利的声望终于开始上升，渐渐成为印度一位受人尊敬的物理学家。印度物理学界忽视了雷查德利那么久，但最终，在雷查德利晚年以及去世之后，毕竟还是他们怀念和记录了他的生平。而西方国家的广义相对论学者虽然对雷查德利方程给予了很大的重视，对雷查德利本人的了解却微乎其微。在霍金等人的广义相对论专著中都介绍并使用了雷查德利方程，但都没有直接提及雷查德利的原始论文，更不用说对他进行片言只语的生平介绍了。作为物理学家的雷查德利完全湮没在了作为方程名称的雷查德利的背后。我们现在所知有关雷查德利的生平资料，几乎全都出自印度物理学家之手（当然，其中不免有一些过誉之语）。在雷查德利去世前，印度校际天文与天体物理中心（Inter-University Centre for Astronomy and Astrophysics）及印度科技部所设的科普组织“科学传播”（Vigyan Prasar——这个英文名称是印地语音译）拍摄了一部以他的生平和成就为内容的纪录片。可惜雷查德利本人没能看到这部纪录片的问世。

雷查德利去世的时间距离他发表雷查德利方程恰好相隔半个世纪。

2004年2月，雷查德利在《广义相对论与引力》（General Relativity and Gravitation）杂志上发表了一篇题为“非旋转理想流体的无奇点宇宙学解”（Singularity-Free Cosmological Solutions with Non-Rotating Perfect Fluids）的文章。这是他一生所写的最后一篇论文，对奇点及相关课题的研究伴随他走完了整个孤独的学术生涯。

注释

[\[1\]](#) 当然，这里所谓的“其他物理理论”指的是不把时空本身作为研究对象的理论。另外，人们通常把由非奇异初始条件演化而来的奇点称为理论本身所具有的奇点，以区别于纯粹通过初始条件而引入的奇点。显然，线性理论——比如麦克斯韦电磁理论——不可能有这样的奇点。

[\[2\]](#) 洛伦兹度规是指符号差（signature）为（1，—1，—1，—1）的度规（有些文献的定义与我们的相差一个整体符号）。除洛伦兹度规外，人们还常常在时空定义中附加一些其他条件，比如豪斯道夫条件（Hausdorff condition）、连通性，等等。对于度规的可微性则有的假定为 C^∞ ，有的假定为 C^r （ r 为正整数——请读者思考一下， r 最小应该是多少？），等等。

[\[3\]](#) 有些物理学家试图将奇点视为时空流形的边界，称为奇异边界（singular boundary），但这方面迄今尚未建立起令人满意的处理方式。

[\[4\]](#) 非类空是类时与类光的统称。这里所说的试验粒子包括零质量粒子。

[\[5\]](#) 一条时空曲线不可延拓，直观地说就是在时空流形内没有端点。为此我们首先要求时空流形本身是“不可延拓”的，即无法等度规地（isometrically）嵌入更大的流形中（不过如何实现这一要求本身就是一个很艰深的问题）。这一要求排除了一些平凡的奇点，比如在闵科夫斯基时空（Minkowski spacetime）中挖去一个时空点所造成的“奇点”。测地线的不可延拓性则可以用来排除诸如施瓦西视界这样的表观奇点。

[\[6\]](#) 我们也可以完全类似地定义类空测地不完备性，但由于沿类空测地线的运动是物理上不可实现的，因此这种测地不完备性在奇点研究中不如其他两种测地不完备性那样受重视。

[\[7\]](#) 这个例子比较平凡，一个更复杂的例子是所谓的Taub-NUT空间，它具有 $R^1 \times S^3$ 拓扑结构，曲率张量处处有界，但同样是测地不完备的（类时与类光都不完备）。

[\[8\]](#) 这个例子比较特殊，一个更具物理意义的例子是雷斯勒-诺斯特朗姆（Reissner-Nordström）解，它描述的是带质量及电荷的球对称时空，雷斯勒-诺斯特朗姆解具有类光测地完备性，但不具有类时测地不完备性。

[\[9\]](#) 文献中对这一张量的称呼很多，除“涡旋”外，常见的叫法还有“扭变”（twist）与“旋转”（rotation）。

[\[10\]](#) 确切地讲，这是将曲率张量的定义用于测地切矢量场所得到的关系式。

[\[11\]](#) 需要提醒读者注意的是，这里介绍的只是雷查德利方程应用于测地切矢量场的特例，一般情况下雷查德利方程的适用范围并不限于测地切矢量场。

[\[12\]](#) 类时一般性条件也可以表示成这一形式，因此有些文献对两类一般性条件统一使用这一形式。

[\[13\]](#) 从某种意义上讲，一般性条件也是对物质分布的间接限定。这种限定不同于我们在1.2节中介绍的那些能量条件，因为平直时空满足所有那些能量条件，却不满足一般性条件。有鉴于此，有些文献把一般性条件通俗地表述为“时空中存在物质”。不过要注意的是：由于一般性条件涉及的是曲率张量而不是里奇（Ricci）张量，因此不能被简单地约化为对物质能量动量张量的直接限制。

[\[14\]](#) 在某些文献中，强能量条件与一般性条件被合称为一般能量条件（generic energy condition）。

[\[15\]](#) 雷查德利当时研究的是非旋转尘埃物质的运动。在雷查德利之前，奥地利逻辑学家哥德尔（Kurt Gödel, 1906—1978）在1949年发表的有关哥德尔解的著名论文中率先用到了一些在奇点定理的证明中有着重要作用的技巧及思路。这些工作都是奇点定理的先声。

[\[16\]](#) $E^+(S)$ 和 $E^-(S)$ 分别被称为 S 的未来边界（future horismos）和过去边界（past horismos）。

[\[17\]](#) 严格讲，我们还必须证明被选为前提的那4个条件彼此相容（因为5个条件不相容并不保证其中4个一定相容）。对于在奇点定理中假定为前提的条件（2）～（5）来说，这一证明可以通过给出一个满足条件（2）～（5）的广义相对论的解来完成。

[\[18\]](#) 这种非时序点集 S 的边界定义为这样一些点 p 的集合： p 的任何开邻域都包含一个位于 p 的类时过去的点，一个位于 p 的类时未来的点，以及一条连接这两个点但与 S 不相交的类时曲

线。可以证明，这种非时序点集必定是三维子流形。

[\[19\]](#) 在本系列的讨论中我们没有对时空流形的微分结构进行细分。如果细分的话，霍金-彭罗斯奇点定理要求度规张量起码是二阶连续可微（即 C^2 ）的。这一条件对于某些严格解——比如物质密度存在突变的解——来说并不成立。不过对于具有现实物理意义的情形来说，度规张量即物质密度通常都被认为是足够可微的。

[\[20\]](#) 举一个最简单的例子：由 $\neg(A \wedge B)$ （即A和B不能同时成立）既可以推知 $A \rightarrow \neg B$ （即由A成立推出B不成立），也可以推知 $B \rightarrow \neg A$ （即由B成立推出A不成立）。

[\[21\]](#) 当然，这样的定理将不再是奇点定理，而应该被称为时序破坏定理，或时间机器存在定理。

[\[22\]](#) 在这一研究中，梯普勒还假定了物质的能量密度在过去类时曲线上处处大于一个非零正值，所得到的结果则是时空必定是类光测地不完备的（从而也是测地不完备的）。

[\[23\]](#) 有读者可能会问：既然无论时序条件是否被破坏，奇点都会出现，那为什么不干脆把时序条件从奇点定理的前提中去掉呢？这是因为在论证时序条件的破坏会导致测地完备性的破坏时，往往要引进一些额外的条件（比如前一个注释中提到的梯普勒所用的额外条件）。这些看似细微的额外条件的使用，使得我们无法将时序条件从奇点定理的前提中简单地去除。

[\[24\]](#) 奇点的不可避免性也受到诸如施瓦西解等物理上相当合理的解的直接支持。不过需要指出的是，从纯理论的角度讲，奇点定理所要求的许多条件都是可以被突破的，人们也的确因此而构造出了没有奇点的解。只不过那些解都是非常特殊的，对于具有现实意义的解，广义相对论中奇点的出现几乎是不可避免的。

[\[25\]](#) 当时包括爱因斯坦在内的许多人认为施瓦西解在 $r=2m$ 处的“奇点”是不可到达的，雷查德利的论文对这种看法提出了异议。雷查德利去世后，一些印度物理学家对这一工作给予了很高的评价。不过那些评价有些过誉，因为 $r=2m$ 是否可以到达之所以成为问题，在很大程度上是由于当时许多人误将 $r=2m$ 当成奇点（即所谓的“施瓦西奇点”）。而 $r=2m$ 的非奇异性其实早在1933年就由勒梅特等人指出过（参阅1.1节的注释），那才是解决“施瓦西奇点”问题的关键性工作。

[\[26\]](#) 这一工作其实1953年就完成了，花了两年时间才得以发表。

[\[27\]](#) 有读者可能会问：雷查德利比彭罗斯等人早了整整十年就对奇点进行研究，并且得

到了在奇点定理的证明中起到重要作用的雷查德利方程，为何却没能率先证明奇点定理？这其中主要的原因是因为在奇点定理的证明中还需要大量严密而精巧的几何论证，以及对时空因果性质的细致分析。从雷查德利的研究风格来看，那样的论证和分析明显超出了他的数学背景和研究范围。另一方面，雷查德利一直在试图寻找的没有奇点的时空，这种动机也在无形中影响了他进一步努力的方向。

[\[28\]](#) 在科学培训协会期间，雷查德利在其“正业”金属性质研究中也取得了一定的成果。

第3章 正质量定理

3.1 渐近平直时空

自本节开始，我们将介绍一个新的专题——正质量定理（有时也称为正能量定理）。这一定理是经典广义相对论中一个很漂亮的结果，但在广义相对论的教材甚至专著中都极少介绍，读者要了解这一定理，往往只能求助于原始文献。而原始文献与教材或专著的一个很大的区别，就在于它的论述往往不是自给自足（self-contained）的，而要依赖其他原始文献。更麻烦的是，那些被依赖的原始文献本身——由于也是原始文献——又分别有自己所依赖的原始文献……这种层层依赖的结果是，除非读者已有足够的背景知识，否则为读懂一篇原始文献，往往要顺藤摸瓜地读上一大堆其他文献，其过程有如一个初学英语的人试图通过英-英词典查找词义，往往在词义之中又发现生词。对于正质量定理来说，这种困难在舍恩与丘成桐的论文中体现得尤为明显。在他们高度数学化的论文中，对物理背景及某些源自物理的数学表达式的由来交待得极为简略，或基本不做交待。不熟悉背景的读者哪怕细细研读他们的论文，也有可能只见树木，不见森林。因此在进入正题之前，我们将首先对正质量定理及其证明将会涉及的若干物理概念进行介绍。

引力场的能量动量问题一直是广义相对论研究中一个很困难的课题。自爱因斯坦以来，许多早期的广义相对论研究者，比如美国物理学家托曼（Richard Tolman, 1881—1948）、苏联物理学家朗道、丹麦物理学家莫勒（Christian Møller, 1904—1980）等，都曾在这一问题上做过很大努力。那些早期研究的目的之一，是想探究引力场本身的能量动量是如何分布的。现在回顾起来，那些研究虽然绝非毫无启示，但在很

大程度上归于了失败。后来的物理学家们大都认为，引力场的能量动量是不可定域的。从等效原理的角度讲，这一点几乎是显而易见的，因为对应于牛顿引力场的黎曼联络（Riemannian connection）可以通过坐标变换局域地消去。因此如果执意要寻找引力场能量动量分布的定域表述，就得付出很大的代价，比如限制坐标变换，引进高阶导数，放弃能量动量的协变性，设定背景度规或背景联络，等等。这些做法没有一种是省油的灯，而且过去多年的研究经验表明，即便付出那样的代价，结果依然不尽人意。美国物理学家米斯纳（Charles Misner, 1932— ）、索恩（Kip Thorne, 1940— ）和惠勒在其巨著《引力》（Gravitation）中曾经表示：寻找定域引力场能量动量的努力是“试图为一个错误的问题寻找正确的答案”。[\(1\)](#)

但另一方面，一对脉冲双星会因引力辐射而损失能量，从而导致轨道蜕变及轨道周期的变化，这种周期变化可以精确地加以计算，并获得观测的验证[\(2\)](#)。由于引力辐射是由纯引力场组成的，因此引力场本身携带能量动量是毫无疑问的事情。这样看来，问题的关键在于有关引力场的能量动量我们究竟可以知道多少，这是一个目前还在研究之中的问题。物理学界比较公认的一点是，一个孤立体系的总能量动量是可以定义的，这个总能量动量既包含了普通物质的贡献，也包含了引力场的贡献。对这一能量动量的理论表述是由美国物理学家阿诺维特（Richard Arnowitt, 1928— ）、戴舍（Stanley Deser, 1931— ）和米斯纳于20世纪60年代初提出的，被称为**ADM**能量动量，其中的能量部分则称为**ADM**能量（ADM energy），也叫**ADM**质量（ADM mass）。我们所要介绍的正质量定理涉及的就是**ADM**质量[\(3\)](#)。

要想定义一个孤立体系的总能量动量，首先必须搞明白什么是孤立体系。从物理上讲，一个体系的“孤立”指的是远离任何其他体系，或者更确切地说，是任何其他体系对它的影响都可以忽略。对于这样的体系，我们可以在各种距离上考察其物理性质，而不必担心受到其他体系

的影响。由于这一特点，我们可以利用该体系的某些与距离有关的物理性质来定义其“孤立”性。比方说，我们可以这样来定义一个孤立的带电体系：

- (1) 所有的源（电荷、电流）都分布在一个紧致的空间区域中；
- (2) 对于固定时刻 t ，沿空间方向远离体系——即空间距离 $r \rightarrow \infty$ ——时场强的衰减与空间距离的平方成反比，即 $F_{\mu\nu} \sim O(1/r^2)$ ；
- (3) 沿类光测地线远离体系时场强的衰减与空间距离成反比，即 $F_{\mu\nu} \sim O(1/r)$ 。（请读者想一想，这一渐近行为的物理意义是什么？）

将这种思路运用到引力场中，一个很自然的设想是通过度规场 $g_{\mu\nu}$ 以适当方式趋于闵科夫斯基度规 $\eta_{\mu\nu}$ 来定义孤立体系。因为从物理上讲，远离一个孤立体系时，时空应该是平直的，这样的时空被称为渐近平直时空（asymptotically flat spacetime）。这种定义实质上是想通过时空的渐近平直性来定义孤立体系，这是现代广义相对论研究所采用的方法。在早期研究中，人们通常将渐近平直时空定义为度规满足条件

$$\begin{cases} g_{\mu\nu} \sim \eta_{\mu\nu} + O\left(\frac{1}{r}\right) \\ \partial_i g_{\mu\nu} \sim O\left(\frac{1}{r^2}\right) \end{cases} \quad (3.1.1)$$

的时空（视情形不同，有时还会附加对 $g_{\mu\nu}$ 更高阶导数的限制）。我们将会看到，这一定义所要求的渐近行为并非随意选取，而是出于确定引力场总能量动量的需要。不过，这一定义——我们称之为朴素定义——虽然在不少具体场合是适用的，对于普遍研究来说却有很大的局限性。这首先是因为它明显依赖于坐标的选择，比如定义中所用的空间距离 r 就是一个依赖于坐标选择的概念。在特殊的坐标下， $r \rightarrow \infty$ 甚至有可能只是一个有限远的点。不仅如此，用度规场在 $r \rightarrow \infty$ 的行为来定义渐近平直

时空还有一个技术上的不利之处，那就是处理 $r \rightarrow \infty$ 的极限并不是一件容易的事情，尤其是在不得不涉及多个极限或微分的相互次序的时候。

为了克服上述困难，自20世纪60年代起，彭罗斯等人提出了一个非常聪明的想法，那就是干脆把 $r \rightarrow \infty$ （即“无穷远”）这个“麻烦制造者”当作边界加入到时空流形中。这样一来，度规的渐近行为就可以用其在流形边界及其邻域内的微分性质来取代，从而避免采用像 $r \rightarrow \infty$ 这样依赖于坐标选择的极限。当然，这个想法说起来容易，具体做起来却有不少微妙的细节需要处理。彭罗斯最初的工作只考虑了对类光无穷远的处理，在他之后，经过杰罗奇、美国加州大学圣塔芭芭拉分校（University of California, Santa Barbara）的物理学家霍罗威茨（Gary Horowitz）、印度物理学家阿什提卡（Abhay Ashtekar, 1949— ）等多位物理学家的努力，直到20世纪七八十年代，人们才得到了渐近平直时空的相对完整的现代定义。

虽然我们所要介绍的正质量定理的表述和证明只需用到渐近平直时空的朴素定义，但我们仍将对现代定义的思路做一个简单介绍。这不仅是因为这一思路本身值得了解，而且也是因为通过对现代定义的介绍，我们可以对朴素定义的适用性有一个了解。如上所述，现代定义的关键是将“无穷远”纳入时空边界，要做到这一点，第一步显然是要对时空进行“压缩”，以便把看不见摸不着的“无穷远”拉到看得见摸得着的“有限远”。而要想对时空进行“压缩”，就必须改变时空的尺度。在数学上，这是通过所谓的共形变换（conformal transformation） $\bar{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ 来实现的。这种变换之所以被称为共形变换，是因为（请读者自行证明）它只改变尺度而不改变角度，从而也不改变几何形状，其中包括光锥的形状。在相对论中，时空的因果结构是由光锥决定的，因此共形变换不改变时空的因果结构，这是它在广义相对论研究中广受青睐的根本原因。经过适当的共形变换（请读者想一想，为了将“无穷远”拉到“有限

远”，共形因子 Ω 需要满足的最基本的条件是什么？），再辅以一定的坐标变换，时空流形可以最终用一组在有限区间内取值的坐标来描述，这样就完成了将“无穷远”拉到“有限远”的任务。

完成了这一步之后，我们就可以为时空流形添加边界，那些边界点表示的就是原先可望不可及的“无穷远”，在这里被称为共形无穷远（conformal infinity），它们可分为：过去（未来）类时无穷 i^- （ i^+ ），过去（未来）类光无穷 j^- （ j^+ ），以及类空无穷 i^0 。在时空流形上添加边界所得到的结果有时被称为“非物理时空”（unphysical spacetime）[\(4\)](#)，得到这一“非物理时空”的过程则被称为共形紧致化（conformal compactification）。熟悉复变函数论的读者可能注意到了，对时空流形的这种处理方式类似于复变函数论中引进黎曼球面及无穷远点的做法。

为了更好地理解共形紧致化，我们来看一个简单的例子。我们知道，闵科夫斯基度规（在有关正质量定理的这一专题中，我们将采取一个与本书其余部分相反的约定，即将度规的空间部分取为正定，以便更方便地表示空间距离及类空超曲面的性质）

$$ds^2 = - dt^2 + dr^2 + r^2 d\Omega^2 \quad (3.1.2)$$

所涉及的坐标 r 和 t 的取值范围都是无界的。但如果我们对这一度规接连实施下列三个变换：

（1）坐标变换 $u=t-r$ ， $v=t+r$

（2）共形变换 $\bar{g}_{\mu\nu} = [4 / (1+u^2)(1+v^2)] g_{\mu\nu}$

（3）坐标变换 $T = \arctan v + \arctan u$ ， $R = \arctan v - \arctan u$

就可将之转变为（请读者自行验证）

$$ds^2 = - dT^2 + dR^2 + \sin^2 R d\Omega^2 \quad (3.1.3)$$

其中 R 和 T 的取值范围满足： $-\pi < T-R < \pi$ ， $-\pi < T+R < \pi$ ， $R \geq 0$ 。这时所有坐标的取值全都成为了有界，这就完成了前面介绍的将“无穷远”拉

到“有限远”的任务。这一度规所对应的共形无穷远（即代表“无穷远”的边界）则是：过去类时无穷远 i^- 为 $R=0$ ， $T=-\pi$ ，未来类时无穷远 i^+ 为 $R=0$ ， $T=\pi$ ，两者均为一个点；过去类光无穷远 j^- 为 $T=R-\pi$ ， $0<R<\pi$ ，未来类光无穷远 j^+ 为 $T=\pi-R$ ， $0<R<\pi$ ，两者均为三维表面（拓扑结构为 $S^2\times R^1$ ）；类空无穷远 i^0 为 $T=0$ ， $R=\pi$ ，是一个点。

闵科夫斯基度规的这个例子虽然简单，但从中我们可以看到普遍定义所需处理的一个微妙的细节，那就是类空无穷远被映射为了一个点 i^0 。我们知道，物理量（尤其是带分量的物理量，比如矢量或张量）沿不同方向趋于类空无穷远所具有的极限往往是不同的（请读者举出一个具体例子），要想让这些不同的极限与单一的类型空无穷远点相对应，“非物理时空”在这一点上的解析性质必须足够弱⁽⁵⁾。此外，在有物质存在的情形下，由于物质不会凭空产生和消失，因此我们不能要求时空在类时无穷远 i^- 和 i^+ 趋于本质上是描述真空的闵科夫斯基时空。除这些细节外，渐近平直时空的定义还需满足以下三个物理上显而易见的条件：

（1）适当的完备性条件，以确保所有的无穷远都被包含在内。

（2）物质只分布在有限的空间区域内。这一条件原则上可以放宽为对物质能量动量张量 $T_{\mu\nu}$ 在边界附近趋于零的方式的限制。

（3）度规张量在类空无穷远的某个邻域内趋于闵科夫斯基度规。这一条件所体现的是渐近平直的核心含义，它可以被巧妙地表示为 Ω 和 $\bar{g}_{\mu\nu}$ 在边界上的约束条件，这一点也正是现代定义的优势所在。

渐近平直时空的现代定义就是由上述所有条件的数学表述组成的。但光有这样的表述还不够，我们还必须证明这一定义是无歧义的。这一点之所以需要证明，是因为满足定义要求的共形因子 Ω 是不唯一的。由于 Ω 的选择直接影响到时空边界（即“无穷远”）的结构，因此 Ω 的不唯一性原则上有可能导致所定义的时空边界（即“无穷远”）出现歧义。幸

运的是，杰罗奇等人早已证明，不同的 Ω （当然它们首先得满足定义所要求的各种解析性质）所对应的“非物理时空”之间必定可以找到在物理时空上为恒等映射的微分同胚。存在这样的微分同胚表明， Ω 的选择不会影响时空的拓扑及微分性质，从而不会导致歧义⁽⁶⁾。这样就确立了渐近平直时空的现代定义。

渐近平直时空——以及以之为基础的孤立体系——的现代定义虽然抽象，并且看上去与早期研究所用的朴素定义很不相同，但上述条件中的第三条意味着，我们总可以找到一组适当的坐标，使度规张量在“远处”（即类空无穷远的某个邻域内）以朴素定义所要求的方式趋于闵科夫斯基度规。这表明，渐近平直时空的朴素定义虽不严格，但在实际应用中的确是能够行之有效的。

3.2 广义相对论的动力学

有了孤立体系的定义，下面我们来讨论孤立体系的总能量动量。由于能量动量是描述体系动力学行为的物理量，因此在定义它之前，有必要对广义相对论的动力学有所了解。我们知道，爱因斯坦场方程的左边包含时空度规 $g_{\mu\nu}$ 及其一、二阶导数，右边则是描述物质分布的能量动量张量 $T_{\mu\nu}$ 。场方程的这种形式导致了一种流传很广的误解，即以为所谓广义相对论的动力学，就是在时空流形上给定 $T_{\mu\nu}$ ，求解 $g_{\mu\nu}$ 。这种说法即便不算完全错误，起码也是似是而非的。因为时空流形是四维的，因此所谓“在时空流形上给定 $T_{\mu\nu}$ ”，实际上包括了给出 $T_{\mu\nu}$ 作为时间的函数。但所谓动力学，它的目的就是寻找物理量——无论其描述的是几何还是物质——随时间的演化，既然如此，又怎能事先就给定 $T_{\mu\nu}$ 作为时间的函数呢？这既没有现实可行性，也不符合动力学的要求。

那么广义相对论的动力学究竟该如何定义呢？我们可以回想一下普通力学。在普通力学中，要解决一个动力学问题，往往需要给定所谓的初始条件——即一组动力学变量及其时间导数在初始时刻的空间分布。对于广义相对论来说，要想给出初始条件，首先要对时空进行某种分解，这样才能谈论所谓的“初始时刻”和“空间分布”。这种分解的实质，是将时空流形分解为 $\Sigma^3 \times \mathbf{R}^1$ ，其中 Σ^3 （下文将省略维数上标）是三维类空超曲面，坐标记为 x_i （ $i=1, 2, 3$ ，下同），表示空间； \mathbf{R}^1 则表示时间，坐标记为 t 。这种分解通常被称为3+1分解，或**ADM**分解

（Arnowitt-Deser-Misner decomposition）。在这样的分解下，每个时刻 t 都有一个对应的类空超曲面 Σ 。所谓初始条件，就是在某个给定时刻 t_0 （即所谓“初始时刻”），给出一组动力学变量及其时间导数在与该时刻相对应的类空超曲面 Σ 上的数值（即所谓“空间分布”）。

有了时空流形的分解，接下来就可以定义动力学变量。为此，我们

引进 Σ 的单位法矢量 \mathbf{n} ，以及度规 $g_{\mu\nu}$ 在 Σ 上的诱导度规 $h_{ij} = g_{ij} + n_i n_j$ 。细心的读者可能注意到了，诱导度规 h_{ij} 其实就是2.2节中引进的时空度规的“空间部分”（不过两者之间有一个符号差异，请读者想一想，这个差异从何而来？）。 h_{ij} 作为 Σ 上的诱导度规，显然是动力学变量， h_{ij} 在 Σ 上的分布则是初始条件的一部分。但光有 h_{ij} 的分布还不够，爱因斯坦场方程和大多数其他动力学方程一样，是二阶微分方程，因此我们还需要知道 h_{ij} 对时间的导数在 Σ 上的分布。不过由于ADM分解中的时间轴通常不与 Σ 正交，使用起来并不方便，因此人们往往用 h_{ij} 沿 Σ 法向 \mathbf{n} 的李导数（Lie derivative），即

$$K_{ij} = \frac{1}{2} L_n h_{ij} \quad (3.2.1)$$

来取代时间导数。这样定义的 K_{ij} 有什么好处呢？首先，它具有良好的坐标变换性质，这是由李导数的性质保证的；其次，它包含了 h_{ij} 的时间演化信息，从而确实可以取代 h_{ij} 对时间的导数来作为初始条件的一部分。事实上，时间基矢 ∂_t 可以在 \mathbf{n} 及（ Σ 上的）空间基矢 ∂_i 组成的坐标系中分解为 $\partial_t = N\mathbf{n} + N^i \partial_i$ （其中 N 和 N^i 为分解系数），而 K_{ij} 则可以表示为

$$K_{ij} = \frac{1}{2N} (\partial_t h_{ij} - N_{ilj} - N_{jli}) \quad (3.2.2)$$

其中 $|i$ 和 j 表示相对于诱导度规 h_{ij} 的协变导数（下同）。从这一结果可以清楚地看到， K_{ij} 包含了 h_{ij} 的时间演化信息（如果时间轴恰好沿 \mathbf{n} 方向，那么 K_{ij} 将正比于 h_{ij} 的时间导数）。最后，但并非最不重要的，是 K_{ij} 具有清晰的几何意义。事实上，它是超曲面 Σ 的外曲率（extrinsic curvature），也叫做第二基本形式（second fundamental form），它描述的是 Σ 在它所嵌入的外部时空流形中的弯曲方式^{[\(7\)](#)}。

为了更好地显示 K_{ij} 作为外曲率的几何意义，同时也为后文介绍舍恩与丘成桐的证明埋个伏笔，我们对 K_{ij} 再稍做一点考察。我们注意到， $K_{ij} = (1/2) L_n h_{ij} = (1/2) L_n g_{ij} = n_{(i; j)}$ ，其中 $n_{(i; j)} = (1/2) (n_{i; j} + n_{j; i})$ 。

$+n_{j; i}$) 是 $n_{i; j}$ 的对称部分。不难看到, 经过这样改写的 K_{ij} 与 2.2 节中测地线束形变的对称部分完全类似。事实上, 我们这里所介绍的内容与那里有关测地线束的描述之间存在明显的对应关系: 测地线的切矢 V 对应于 n , 与 V 垂直的子空间对应于 Σ , 形变对应于 $n_{i; j}$, 形变的对称部分则对应于 K_{ij} 。在 2.3 节中, 我们提到过这样一个结果, 即形变的反对称部分 (即涡旋张量) 在测地线束为超曲面垂直时为零。由于我们这里讨论的单位法矢量 n 和超曲面 Σ 显然满足超曲面垂直关系, 因此形变 $n_{i; j}$ 的反对称部分 $n_{[i; j]}$ 为零。既然 $n_{i; j}$ 的反对称部分为零, 它就完全由对称部分所组成, 而后者正是 K_{ij} 。这就表明

$$K_{ij} = n_{i; j} \quad (3.2.3)$$

这个结果的几何意义非常明确, 它表明一个曲面 (或超曲面) 的外曲率所描述的是单位法矢量沿曲面运动时的变化, 这完全符合我们从曲面所嵌入的外部空间 (或时空) 来判断其弯曲程度的直觉。

现在我们回到广义相对论的动力学上来。从上面的分析可以看到, 广义相对论时空部分的动力学变量可以选为诱导度规 $h_{ij}(t, x^i)$, 相应的初始条件则是初始时刻所对应的类空超曲面 Σ 上的诱导度规 $h_{ij}(x^i)$ 及外曲率 $K_{ij}(x^i)$ [\[8\]](#)。从动力学变量与方程的数目对比来看, h_{ij} 作为动力学变量是合理的, 因为 h_{ij} 有 6 个独立分量, 而爱因斯坦场方程虽然有 10 个方程, 其中却只有 6 个是动力学方程, 其余 4 个, 即

$$G_{\mu\nu}n^\nu = 8\pi T_{\mu\nu}n^\nu \quad (3.2.4)$$

不包含度规对时间的二阶导数, 因而只是对初始数据的约束条件。这些约束条件可以通过微分几何中所谓的高斯-柯达西方程组 (Gauss-Codacci equations) 表述为有关 h_{ij} 和 K_{ij} 的条件:

$$\begin{cases} {}^{(3)}R + (\text{tr}K)^2 - \text{tr}(K^2) = 16\pi\rho \\ [K_{ij} - (\text{tr}K)h_{ij}]^{lj} = 8\pi J_i \end{cases} \quad (3.2.5)$$

其中⁽³⁾ R 是超曲面 Σ 上的曲率标量； tr 是 Σ 上的迹； $\rho = T_{\mu\nu}n^\mu n^\nu$ 和 $J_i = h_i^\mu T_{\mu\nu}n^\nu$ 则是物质能流密度的时间和空间分量。这是两个很重要的关系式，在舍恩与丘成桐的证明中将会被用到。

至此，我们就完成了对广义相对论的动力学变量及初始条件的介绍。所谓广义相对论动力学中的初值问题，指的就是给定一个类空超曲面 Σ 及其上满足约束条件(3.2.5)的 h_{ij} 、 K_{ij} 及物质分布，求解 h_{ij} 和 K_{ij} 在整个时空流形上的分布。对于我们所讨论的孤立体系来说，我们还要求 Σ 是渐近平直类空超曲面。要做到这一点， Σ 必须经过类空无穷远点 i^0 ，并满足一定的解析条件⁽⁹⁾。在早期的工作中，类空超曲面 Σ 的渐近平直性与时空本身的渐近平直性一样，是通过朴素方式——即对 h_{ij} 、 K_{ij} 及其若干阶导数的渐近行为加以界定——来定义的。除此之外，人们往往还在时空流形上附加一定的因果条件，比如全局双曲条件（其定义可参阅2.4节），这些我们就不讨论了。

3.3 ADM能量动量

接下来我们介绍广义相对论的哈密顿表述（Hamiltonian formulation），ADM能量动量的原始定义就源于这一表述。我们知道，引力场的作用量密度是 $L = (1/16\pi) (-g)^{1/2} R$ ，其中 $g = \det(g_{\mu\nu})$ 是度规张量的行列式。由于（请读者自行证明） $(-g)^{1/2} = Nh^{1/2}$ ，其中 $h = \det(h_{ij})$ 是诱导度规的行列式，以及（可以由高斯-柯达西方程组得到） $R = {}^{(3)}R + \text{tr}(K^2) - (\text{tr}K)^2$ ，上述作用量密度可以表述为 h_{ij} 和 K_{ij} 的函数：

$$L = \frac{1}{16\pi} Nh^{1/2} [{}^{(3)}R + \text{tr}(K^2) - (\text{tr}K)^2] \quad (3.3.1)$$

由此可知与 h_{ij} 对应的广义动量为

$$\pi^{ij} = \frac{\delta L}{\delta h_{ij}} = \frac{1}{16\pi} h^{1/2} [K^{ij} - (\text{tr}K) h^{ij}] \quad (3.3.2)$$

而引力场的作用量密度则可以改写为（其中丢弃了一些全微分项）

$$L = \pi^{ij} \partial_t h_{ij} - NR^0 - N_i R^i \quad (3.3.3)$$

其中

$$\begin{cases} R^0 = -\frac{1}{16\pi} h^{1/2} \left[{}^{(3)}R - \frac{\text{tr}(K^2)}{h} + \frac{(\text{tr}K)^2}{2h} \right] \\ R^i = -\frac{1}{8\pi} \pi^{ij} |_{,j} \end{cases} \quad (3.3.4)$$

从这里我们可以看到，引力理论是一个有约束的理论，约束条件为 $R^0 = 0$ 和 $R^i = 0$ ， N 和 N^i 是相应的拉格朗日乘子。

由上述作用量密度可得引力场的哈密顿密度为

$$H = \pi^{ij} \partial_t h_{ij} - L = NR^0 + N_i R^i \quad (3.3.5)$$

相应的哈密顿量（Hamiltonian）则为 $\int H dV$ （积分为类空超曲面 Σ 上的体积分）。至此，似乎一切都很顺利。但不幸的是，这个哈密顿量——作为 h_{ij} 和 K_{ij} 的泛函——可以描述引力场的动力学，却无法用来计算总能量，因为约束条件 $R^0=0$ 和 $R^i=0$ 的存在使它的数值恒等于零。要解决这个问题，必须注意到引力理论的一个微妙的细节，那就是在哈密顿量的变分中，有一些边界项在 Σ 渐近平直的情况下不为零。比如假定 Σ 的法矢量 n 在类空无穷远处渐近于时间平移生成元 $\partial/\partial t$ ，即 $N \rightarrow 1$ ， $N^i \rightarrow 0$ ，那么可以证明，哈密顿量的变分中有这样一个非零的边界项（它也是唯一一个非零的边界项）：

$$- \int \frac{1}{16\pi} N h^{1/2} h^{ij} \delta^{(3)} R_{ij} dV \rightarrow - \delta \int \frac{1}{16\pi} (\partial^i h_{ij} - \partial_j h_i^i) dS^j \quad (3.3.6)$$

其中 dS^j 是边界 $\partial\Sigma$ 上的面积元，积分理解为 $\partial\Sigma$ 趋于 i^0 的极限（下同）。显然，如果我们在哈密顿量上添加一个相应的边界项 $(1/16\pi) \int (\partial^i h_{ij} - \partial_j h_i^i) dS^j$ ，就可以抵消变分中的边界项。这一附加的边界项对动力学方程没有影响，却对总能量有贡献。事实上，由于哈密顿密度的体积分分为零，这一边界项是对总能量的唯一贡献。这样，我们就得到了孤立体系的总能量 [\(10\)](#)

$$E = \frac{1}{16\pi} \int (\partial^i h_{ij} - \partial_j h_i^i) dS^j \quad (3.3.7)$$

类似地，通过对 N 及 N^i 的渐近行为做其他假设，可以得到孤立体系总动量的表述：

$$p^i = \frac{1}{8\pi} \int (K_j^i - (\text{tr} K) \delta_j^i) dS^j \quad (3.3.8)$$

可以证明，这样定义的 E 和 p^i 与 Σ 上渐近坐标的选择无关，并且构成类空

无穷远 i^0 上的四维矢量。这就是所谓孤立体系的**ADM**能量动量，其中的 E 被称为ADM能量或ADM质量， p^i 被称为ADM动量。由上述定义不难看到，ADM能量动量的存在要求 $\partial_k h_{ij} \sim O(1/r^2)$ ，而这正是渐近平直时空的朴素定义所要求的。

3.4 正质量定理

在上一节中，我们定义了ADM质量。在最简单的情况下，ADM质量是很容易计算的。作为练习，读者不妨计算一下施瓦西时空的ADM质量，以验证它与施瓦西度规中质量参数 m 的等同性。自ADM质量被提出以来，物理学家们一直有一个猜测，那就是ADM质量作为一个孤立体系的总能量，应该是非负的。这个猜测被称为正质量猜想。20世纪70年代末80年代初，正质量猜想被舍恩（曾经是丘成桐的学生）、丘成桐及美国数学物理学家威顿（Edward Witten, 1951— ）所证明，从而成为了正质量定理。

正质量定理：一个孤立体系若其物质分布满足主能量条件，则其ADM质量非负。

在正质量定理中，孤立体系如我们在3.1节中所述，是通过渐近平直时空来定义的。为了使ADM质量本身有定义，正质量定理隐含了一个条件，那就是时空中存在具有良好解析性质的渐近平直三维类空超曲面 Σ 。正质量定理是一个有很大普适性的定理，这普适性主要体现在两个方面：一是物质分布只需满足主能量条件，而无须附加更精细的限制；二是 Σ 的拓扑结构可以有很大的复杂性，比如它可以包含不止一个渐近平直区域，并且各个渐近平直区域中的ADM质量可以彼此不同（这种时空的一个例子是我们以后将要介绍的虫洞）。另外要说明的是，所谓ADM质量非负，更“协变”地讲是指ADM能量动量非类空。

在文献中，正质量定理还包含另外一层含义（我们将不展开讨论），那就是如果ADM质量为零（在存在不止一个渐近平直区域的情形下，则如果ADM质量在其中一个渐近平直区域为零），那么 Σ 必定是三维欧几里德空间，而时空本身则必定是闵科夫斯基时空。这表明，闵

科夫斯基时空不仅是所有时空中能量最低的，而且是具有最低能量值的唯一时空。这一点对于确保闵科夫斯基时空的稳定性有很重要的意义⁽¹¹⁾，同时它也意味着所有非平凡渐近平直时空的ADM质量都是正的，这是正质量定理中“正”字的含义所在。

正质量定理的证明花费了数学家和物理学家们整整20年的时间，它的证明者丘成桐和威顿由于其在证明这一定理及若干其他领域中的卓越贡献，于1982年和1990年先后获得了有“数学诺贝尔奖”之称的菲尔茨奖（Fields Medal）。这些事实都说明，正质量定理是一个艰深的数学问题，它的证明是第一流的数学成就。但另一方面，从物理上讲，一个孤立体系——尤其是经典意义下的孤立体系——的总能量非负难道不应该是显而易见的结果吗？事实上，在经典物理学的任何其他分支中，都从未出现过如此艰深的正质量定理，却唯独在广义相对论中出现了例外，这是为什么呢？定性地讲，一个孤立体系——哪怕是经典意义下的孤立体系——的总能量非负之所以不像想象的那样显而易见，是因为物质之间的引力结合能是负的。在牛顿引力理论中，一个孤立点质量产生的引力场的能量不仅是负的，而且可以趋于负无穷，这样的孤立体系的总能量也完全可以负的⁽¹²⁾。由此可见，类似于正质量定理这样的结果在牛顿引力理论中是不存在的，孤立体系的总能量非负非但不是显而易见的结果，而且还是广义相对论有别于牛顿引力理论的一个重要特征（请读者想一想，电磁理论中也存在负的相互作用能，为什么却没有造成同样的困难？）。

细心的读者也许会提出这样一个问题：我们在本节一开始曾经提到，施瓦西时空的ADM质量正好等于施瓦西度规的质量参数 m 。如果我们将这个质量参数取为负的，即 $m < 0$ ，不就可以得到一个负的ADM质量了吗？从单纯的数学计算上讲情况的确如此。但是， $m < 0$ 与 $m \geq 0$ 的施瓦西时空之间存在着一些本质差异。对于我们所讨论的ADM质量来说，最重要的差异在于：在 $m < 0$ 的施瓦西时空中，所有渐近平直三维

类空超曲面都会经过奇点 $r=0$ ，从而破坏ADM质量的定义（ $m \geq 0$ 的施瓦西时空中则存在不经过奇点的渐近平直三维类空超曲面，从而使ADM质量有良好的定义）[\(13\)](#)。此外，我们在3.2节的末尾曾经提到，在研究广义相对论动力学的时候，人们往往在时空流形上附加一定的因果条件，比如全局双曲条件。 $m < 0$ 的施瓦西时空很不幸地会破坏这类条件。从这个意义上讲，它的动力学本身就缺乏良好的定义，更遑论ADM质量。如果要追根溯源的话，那么这一切的困难都是由奇点 $r=0$ 造成的。对于 $m < 0$ 的施瓦西时空来说，这一奇点是一种很特殊的奇点，被称为裸奇点，这是一种人见人怕（甚至连上帝都怕）的东西，我们将会在下一个专题——宇宙监督假设——中加以讨论。有读者可能会进一步问：既然问题出在奇点上，那么如果我们将点源换成非奇异的物质分布，是否就可以绕开这些问题呢？答案是否定的，因为 $m < 0$ 的施瓦西时空无法与任何满足主能量条件的物质分布相匹配（从这里我们可以看到能量条件对于正质量定理的意义）。因此， $m < 0$ 的施瓦西时空并不满足正质量定理所要求的物理条件，从而不构成针对这一定理的反例。

就像数学或物理中几乎所有的困难问题一样，正质量定理在其最终证明出现之前，就曾经有数学家和物理学家给出过不完整或特殊情形下的证明。这方面最早的工作甚至出现在有关ADM质量的主要论文发表之前。早在1959年，美国普林斯顿大学（Princeton University）当时还刚刚成为博士的物理学家布里尔（Dieter Brill）就在 Σ 上的初始条件轴对称的情形下，对一类特定的引力波进行了研究（这类引力波有时被后人称为布里尔波），并证明了其能量非负。1968年，布里尔与戴舍

（即“ADM”中的“D”）合作证明了当 Σ 上的初始条件与平坦条件（即闵科夫斯基时空下的初始条件）之差的二阶以上效应可以忽略时，正质量定理成立（这一结果正是我们在前面注释中提到的结果，即闵科夫斯基时空起码是准稳定的）。20世纪70年代，雷波维茨（Clement Leibovitz）、伊斯雷尔（Werner Israel, 1931— ）、米斯纳

（即“ADM”中的“M”）等人先后证明了在 Σ 上的初始条件球对称的情形下，正质量定理成立。1976年，法国数学家肖盖（Yvonne Choquet-Bruhat, 1923— ）和加拿大数学家马斯登（Jerrold Marsden, 1942—2010）对布里尔和戴舍1968年的结果进行了推广，证明了当 Σ 上的初始条件与平坦条件之差在某种特定的泛函分析意义上“足够小”（该条件比布里尔和戴舍所要求的“二阶以上效应可以忽略”更弱）时，正质量定理成立。而1977年，莱特（Maria Leite）则给出了 Σ 可以等度规嵌入 R^4 这一特殊情况下正质量定理的证明。上面这些研究都是在纯广义相对论的范围内进行的，其中比较数学化的工作都侧重于从几何角度进行分析，在这一方向上的集大成者是舍恩与丘成桐，他们最终给出了正质量定理的第一种完整证明。

除上述结果外，戴舍和智利物理学家泰特伯姆（Claudio Teitelboim, 1947— ）于1977年在当时还很新颖的超引力（supergravity）理论中发现了一个对于正质量定理的研究有着重要意义的结果，那就是超引力理论的ADM质量非负。这个结果对于广义相对论研究者来说很有些出乎意料，因为超引力理论从总体上讲是一个远比广义相对论复杂的理论，但就ADM质量非负这一特定结果而言，其证明从表观上看，却很奇妙地要比广义相对论中正质量定理的证明容易得多⁽¹⁴⁾。之所以出现这样的结果，是由于在超引力理论中，能量算符可以表示为在纯广义相对论中不存在的旋量性超荷算符（supercharge）的平方，从而具有形式上的非负性。1978年，格里沙鲁（Marc Grisaru）猜测这一结果可以在适当的极限下过渡为广义相对论中的正质量定理。超引力理论的这一特点——尤其是格里沙鲁的猜测——给了威顿很大的启示，使之于1981年给出了正质量定理的第二种证明。

3.5 舍恩与丘成桐的证明概述

从几何角度对正质量定理的研究自1979年起进入了一个大收获的时期。那一年，舍恩与丘成桐发表了一篇长达32页的论文，在假定 Σ 满足一个特定条件 $\text{tr}K=0$ 的基础上证明了正质量定理。

$\text{tr}K=0$ 这一条件在正质量定理的研究中是很常用的。我们在上节中提到的很多工作，从布里尔在20世纪50年代末的工作，到布里尔和戴舍、肖盖和马斯登在20世纪60—70年代的工作，都用到了这一条件。我们首先来看一看，这个条件的几何意义是什么？

读者也许还记得，我们在3.2节中介绍 K_{ij} 的几何意义时，曾表示要“为后文介绍舍恩与丘成桐的证明埋个伏笔”。这个“伏笔”是什么呢？它就是我们在那里介绍的 K_{ij} 与测地线束形变之间的相似性，以及 $K_{ij}=n_{i;j}$ 这一结果。从这些结果立刻可以看到， $\text{tr}K$ ——有时被称为平均外曲率（mean extrinsic curvature）⁽¹⁵⁾——作为 $n_{i;j}$ 的迹，对应的是测地线束形变中描述测地线束会聚或发散趋势的膨胀标量 θ （参阅2.2节），而 $\text{tr}K=0$ 表示的显然是测地线束既不会聚也不发散，换句话说，其体积取极值。这个类比表明， Σ （处处）满足 $\text{tr}K=0$ 这一条件意味着三维类空超曲面 Σ 的体积取极值。进一步的研究表明，这一极值是极小值。在微分几何中，人们将（处处）满足 $\text{tr}K=0$ 的曲面称为极小曲面，但在广义相对论的研究中，它往往被称为极大超曲面（请读者想一想，为什么会出现这样的命名差异？）。因此，被舍恩、丘成桐以及其他研究者所广泛采用的这一条件可以表述为：时空中存在渐近平直的三维类空极大超曲面。

由于舍恩和丘成桐的证明是几何证明，因此，他们需要将主能量条件表述为几何条件。这可以利用我们在3.2节中介绍过的式（3.2.5），即

$$\begin{cases} {}^{(3)}R + (\text{tr}K)^2 - \text{tr}(K^2) = 16\pi\rho \\ [K_{ij} - (\text{tr}K)h_{ij}]^{1j} = 8\pi J_i \end{cases} \quad (3.5.1)$$

来做到。这两个关系式是舍恩和丘成桐的论文的起点，也是本章前两节为读者理解舍恩和丘成桐的论文所做的主要铺垫。利用这两个关系式，主能量条件所要求的物质能流密度矢量非类空 $\rho \geq |J^i J_i|^{1/2}$ 可以被表示为有关几何结构 h_{ij} 和 K_{ij} 的约束条件。由于主能量条件意味着物质能量密度 $\rho \geq 0$ ，在舍恩和丘成桐所考虑的 $\text{tr}K=0$ 的情形下，这显然只有在 ${}^{(3)}R \geq 0$ 的情况下才能满足（请读者自行证明这一点）。因此在 $\text{tr}K=0$ 的情形下，主能量条件要求 ${}^{(3)}R \geq 0$ ，这个几何条件在正质量定理的几何研究中有重要作用，在舍恩和丘成桐的证明中也被反复用到。

舍恩和丘成桐的证明采用的是反证法的思路，即通过假定ADM质量小于零来推出矛盾，其过程大致分为三步⁽¹⁶⁾：首先，他们证明了如果ADM质量小于零，那么在 Σ 中可以构造出一个特殊的二维极小曲面 S ，它在一个紧致集之外满足 ${}^{(3)}R \geq 0$ 。在这一步中，他们用到的是 Σ 渐近平直这一特点，以及 ${}^{(3)}R \geq 0$ 这一来自主能量条件的推论。由于 S 是极小曲面，因此 S 的面积泛函的二次变分必定非负，利用这一点，舍恩和丘成桐——作为第二步——证明了 S 的高斯曲率 K 在曲面上的积分 $\int K dS > 0$ 。在这一步中，他们再次用到了 ${}^{(3)}R \geq 0$ 这一几何条件，以及第一步所得到的在 S 上的一个紧致集之外 ${}^{(3)}R \geq 0$ 这一构造性质。最后，为了推出矛盾，舍恩和丘成桐用两种不同的方法——其中只用到了 Σ 的渐近平直性以及 S 的构造性质——证明了一个与 $\int K dS > 0$ 完全相反的结果，即 $\int K dS \leq 0$ 。这一矛盾的出现表明ADM质量小于零这一假设与证明过程中所用的其他假设不相容。由于证明过程中所用的其他假设都是正质量定理本身的假设（比如 Σ 的渐近平直性）或其推论（比如 ${}^{(3)}R \geq 0$ ），因此这一矛盾的出现表明在正质量定理所假设的条件下，ADM质量必须非负。这样舍恩和丘成桐就完成了在 $\text{tr}K=0$ 这一特定条件下正质量定理的

证明。

舍恩和丘成桐的这一证明无疑是一个很重要的结果，但由于它依赖于 $\text{tr}K=0$ ，或者说 Σ 是极大超曲面这一额外条件，因此与正质量定理的完整证明还有距离。如我们在前面所述， $\text{tr}K=0$ 这一条件在正质量定理的早期研究中是经常用到的，在某些研究中人们几乎已将之视为正质量定理的前提之一。在舍恩和丘成桐的证明问世前不久，肖盖等人曾对这一条件进行了研究，确立了它在一系列合理情形下的有效性。但是，这充其量只能说明这一条件具有一定的合理性，却不足以确立其普遍有效性。要想得到一个真正普遍的正质量定理，这一条件必须去除。但这一条件一旦去除，从主能量条件中就无法得到⁽³⁾ $R \geq 0$ 这一在正质量定理的几何研究中被反复用到的几何条件了（请读者自己证明这一点），从而上面所介绍的舍恩和丘成桐的证明，以及他们之前某些其他人的工作，就都不复成立了。因此这看似一步之差的细节，实际上是牵一发而动全身，其修补难度是非常大的。不过舍恩和丘成桐在发表1979年的论文时，已经对整个证明有了全面构想。他们在发表那篇论文的同一年还发表了一篇摘要性的短文，概述了在更普遍（即不要求 $\text{tr}K=0$ ）的情况下证明正质量定理的思路。在正质量定理被完全证明后，人们有时会回溯到这一年，即1979年，将之作为正质量定理被证明的年份。但事实上，舍恩和丘成桐的证明细节直到两年后的1981年才正式发表。在同一年早些时候，他们还发表了另一篇短文，将正质量定理的表述由ADM质量非负推广为ADM能量动量非类空。

舍恩和丘成桐的完整证明发表后的第二年，曾在相关领域中做过研究的美国宾州大学（University of Pennsylvania）的微分方程及微分几何学家卡斯丹（Jerry Kazdan）撰写了一篇介绍正质量定理的文章，对于舍恩和丘成桐1981年的证明，卡斯丹写下了一句简短的评语：“任何充分的概述都将太长。”现在回过头来品味四分之一世纪前的这句评语，我们感觉到情况并未发生太大变化，舍恩和丘成桐的证明在很大程

度上维持了当初的复杂性。因此我们将接受卡斯丹的忠告，不试图对它做任何“充分的概述”。舍恩和丘成桐在1981年的论文中所做的，用最简单——从而很不充分——的话来概述的话，是证明了在保持ADM质量不变的情况下，可以将普遍情形下的初始数据集 (Σ, h_{ij}, K_{ij}) 变形为满足 $\text{tr}K=0$ 这一条件的初始数据集。做到了这一点，普遍情形就被转化成了已经被证明的 $\text{tr}K=0$ 的情形。这样，舍恩和丘成桐就完成了正质量定理的证明。

3.6 威顿的证明概述

1981年对于正质量定理的证明来说是一个决胜之年。与舍恩和丘成桐发表完整的证明几乎同时，美国普林斯顿大学（Princeton University）的数学物理学家威顿也完成了一篇文章，给出了正质量定理的一种全新的证明方法。我们在3.4节中曾经提到过，威顿的证明受到了超引力理论中ADM质量非负这一结果的启发。超引力理论的结构相当复杂，它与广义相对论的最大差别，乃是除引力场外，还引进了作为引力子超对称伙伴（supersymmetric partner）的引力微子（gravitino）的场。由于引力子是一种玻色子（boson），而超对称理论中玻色子的超对称伙伴是费米子（fermion），因此引力微子是一种费米子，它的场是一种旋量场（spinorfield），它所满足的量子化条件则是反对易的，这意味着其“大小”正比于普朗克常数（Planck's constant）的平方根，从而在普朗克常数趋于零这一经典极限下应该能被忽略。有鉴于此，格里沙鲁——如3.4节所述——提出了一个猜测，那就是超引力理论的ADM质量非负这一结果可以在经典极限下过渡为一个经典广义相对论的结果。显然，这一结果——如果存在的话——应该就是正质量定理。格里沙鲁的这一猜测从未得到过证明，但它对威顿的证明起了很大的启发作用⁽¹⁷⁾。

威顿证明的出发点是 Σ 上的狄拉克方程（Dirac's equation） $\gamma^i D_i \psi = 0$ ⁽¹⁸⁾，其中 γ^i 是狄拉克矩阵（Dirac matrices）， D_i 是四维协变导数在 Σ 上的分量， ψ 是旋量场。威顿着重研究了这一方程的渐近于某个常数旋量 ϵ ——即 $\psi \rightarrow \epsilon + O(1/r)$ 的特殊解（这种特殊解有时被后人称为威顿旋量）。威顿之所以要研究这一狄拉克方程的这类特殊解，是因为这样的解出现在超引力理论中超荷算符的积分表达式中，这也正是超引力理论对威顿的启示所在。

利用 Σ 的渐近平直性，经过不太复杂的推理，威顿证明了对于任何 $\varepsilon \neq 0$ ，上述狄拉克方程必定存在满足 $\psi \rightarrow \varepsilon + O(1/r)$ 的特殊解 ψ ⁽¹⁹⁾。不过，威顿在证明这一存在性时用到了一些并非显而易见的结果，却未予严格论述。类似的不严密性在他的论文中还不止一处，有些甚至可以归为错误。这么多小缺陷同时出现在一篇论文中，对于数学功力极其深厚的威顿来说是颇为罕见的。因此从严密性上讲，威顿的原始证明与舍恩和丘成桐的证明有一定的差距。但幸运的是，威顿的证明发表之后，哈佛大学（Harvard University）的数学物理学家帕克（Thomas Parker）和陶布斯（Clifford Taubes, 1954—）很快就对他的证明做了改进，弥补了那些缺陷。

威顿的证明中另一个关键步骤，是将 Σ 上的狄拉克算符 $\gamma^i D_i$ 与物质的能量动量分布联系起来。这一步的逻辑地位与舍恩和丘成桐的证明中那些将主能量条件表述为几何条件的关系式相类似，其重要性是显而易见的。因为只有建立了那样的联系，才能将主能量条件应用到证明中来。威顿在证明这一步时也出现了疏漏，忽略了狄拉克算符平方展开式中的一个曲率项，不过这一疏漏恰好被他在后文将狄拉克算符的平方展开式作用于 ψ ，并与 ψ 作内积时出现的另一个疏漏所抵消，因此未对整个论证造成实质性的破坏。而且更幸运的是，这一错误在文章付印前就被人发现，使威顿得以及时在文章中增添一个补注加以纠正。经过纠正后的这一联系可以表示为

$$(\gamma^i D_i)^2 = -D^i D_i + 4\pi G(T_{00} + T_{0j} \gamma^0 \gamma^j) + K_i^j \gamma^0 \gamma^i D_j \quad (3.6.1)$$

其中最后一项中的 K_i^j 就是我们在3.2节中引进的 Σ 的外曲率（起初被威顿忽略的正是这一项）。这一结果其实是微分几何中的外森比克公式

（Weitzenböck formula）应用于旋量场的情形。唯一的差别，是利用爱因斯坦场方程对某些几何量进行了替换，从而出现了与物质能量动量张量有关的项。

将上述结果作用于前面提到的特殊解 ψ ，显然可以得到

$$-D^i D_i \psi + 4\pi G(T_{00} + T_{0j}\gamma^0\gamma^j)\psi + K_i^j \gamma^0\gamma^i D_j \psi = 0 \quad (3.6.2)$$

用 ψ 与这一方程作 Σ 上的内积，并经过分部积分（正是在分部积分中，威顿出现了另一个疏漏，恰好抵消了他在狄拉克算符平方展开式中的疏漏），可以得到这样一个结果：

$$\int dS^k \psi^+ D_k \psi = \int dV D^i \psi^+ D_i \psi + 4\pi G \int dV \psi^+ (T_{00} + T_{0j}\gamma^0\gamma^j)\psi \quad (3.6.3)$$

在这一结果中，右端的第一项显然是非负的。而第二项由于 $T_{0j}\gamma^0\gamma^j$ 的本征值为 $\pm\|T_{0j}\|$ ，加上主能量条件要求 $T_{00} \geq \|T_{0j}\|$ ，因此也是非负的。这表明

$$\int dS^k \psi^+ D_k \psi \geq 0 \quad (3.6.4)$$

粗看起来，这个有关旋量场的不等式似乎与正质量定理没什么关系，但威顿注意到它左侧的边界积分是 Σ 上的不变量，而且除 ψ 的渐近值（常数旋量） ε 外，它只与度规张量 h_{ij} 的 $O(1/r)$ 渐近行为有关（更高阶的渐近项对面积分没有贡献）。另一方面，ADM能量 E 和动量 p^i 乃是由 h_{ij} 的 $O(1/r)$ 渐近分量所能构成的仅有的不变量。因此威顿意识到这个不等式左侧的边界积分必定与ADM能量动量有关。那么它们之间究竟是什么样的关系呢？显然需要通过对这一面积分进行计算来揭示。在一般情况下，这将是一个很困难的计算，不过好在 Σ 具有渐近平直性，而面积分又处于渐近平直区域中，这使计算得到了极大的简化。计算的结果给出了一个相当简洁的关系式：

$$\int dS^k \psi^+ D_k \psi = 4\pi G \varepsilon^+ (E + p_i \gamma^0 \gamma^i) \varepsilon \quad (3.6.5)$$

因此前述不等式可以改写成

$$\varepsilon^+ (E + p_i \gamma^0 \gamma^i) \varepsilon \geq 0 \quad (3.6.6)$$

不难看到——与前面运用主能量条件证明 $\int dS^k \psi^+ D_k \psi \geq 0$ 时所用的推理相类似——由于 $p_i \gamma^0 \gamma^i$ 的本征值为 $\pm \|p\|$ ，而且 $\varepsilon \neq 0$ ，因此上述不等式意味着 $E \geq \|p\|$ ，即ADM能量动量非类空。这样，威顿就完成了正质量定理的证明。正质量定理也因此有了两种截然不同的证明。

3.7 讨论

与数学物理中的很多其他定理一样，正质量定理的证明并不代表这一研究方向的终结。事实上，正质量定理被证明之后，很快就有物理学家将注意力转向了它的推广。在本节中我们将对这方面的工作做一个简单介绍。

威顿的证明发表两年后，英国理论物理学家吉本斯（Gary Gibbons, 1946—）等人就将他的证明推广到了渐近平直时空中包含黑洞的情形（舍恩与丘成桐的证明则无须推广就直接适用于吉本斯等人所考虑的情形）。他们并且还证明了，如果所涉及的黑洞含有电荷与磁荷，则ADM质量 $m \geq Q_e^2 + Q_m^2$ （其中 Q_e 和 Q_m 分别为总电荷与总磁荷）。吉本斯等人的证明完全沿用了威顿的方法，只是在协变导数中加入了电磁相互作用项。不过他们的结果有赖于一个附加条件，即电荷密度与磁荷密度的平方和不大于（不包括电磁场的）物质能量密度的平方。稍后，莫里斯奇（Osvaldo Moreschi）和斯帕林（George Sparling）将该结果进一步推广为一个带参数的不等式族。不过，他们的这一工作更多地只是一种纯粹的数学推广，而并无显著的物理意义。

渐近平直时空中包含有黑洞时，一个比吉本斯等人所考虑的更重要的命题是彭罗斯于1973年提出的彭罗斯猜想（Penrose Conjecture），也叫彭罗斯不等式（Penrose inequality），它可以表述为

$$M \geq \left(\frac{A}{16\pi} \right)^{1/2} \quad (3.7.1)$$

其中 M 为ADM质量， A 为所有黑洞的最外部视界面积之和。假如所讨论的黑洞为质量 m 的施瓦西黑洞，则 $A = 16\pi m^2$ ，彭罗斯猜想可以简化为 $M \geq m$ ，即如果渐近平直时空中存在一个质量为 m 的施瓦西黑洞，则整个时空的ADM质量必定不小于 m 。与正质量定理所包含的ADM质量为零

意味着时空为闵科夫斯基时空这层含义相类似，彭罗斯猜想要求等号只在时空为施瓦西时空时才成立。显然，彭罗斯猜想比（包含黑洞的）正质量定理（ $M \geq 0$ ）更强。在后文中我们将会看到，彭罗斯猜想与所谓的宇宙监督假设颇有渊源。

彭罗斯猜想提出二十几年后，2001年，数学物理学家们在证明一类特殊情形（即渐近平直超曲面的曲率标量 $R \geq 0$ 的情形）下的彭罗斯猜想上取得了重要进展。我们在3.5节中介绍舍恩与丘成桐的证明时曾经介绍过 $R \geq 0$ 这一条件，它是 Σ 为极大超曲面情形下主能量条件的推论。数学物理学家们通常把满足这一额外条件的彭罗斯猜想称为黎曼-彭罗斯猜想（Riemannian Penrose conjecture）[\[20\]](#)。1997年，惠斯肯（Gerhard Huisken）和伊尔玛能（Tom Ilmanen）在时空中只包含一个黑洞（确切地说是时空只包含一个单连通视界）的情形下证明了黎曼彭罗斯猜想。两年后（1999年），布雷（Hubert Bray）给出了更普遍情形下黎曼彭罗斯猜想的证明[\[21\]](#)。但迄今为止还没有人能够将 $R \geq 0$ 这一额外条件去除，因此一般情形下的彭罗斯猜想仍未得到证明。

对正质量定理的另一类推广是试图将之推广到四维以上的时空。这也是一个迄今尚未完全解决的问题。在我们介绍过的两种正质量定理的证明中，舍恩和丘成桐的证明往高维方向可以直接推广到八维及八维以下时空，但对于高于八维的时空，他们的方法会遭遇迄今无人能够克服的奇点困扰[\[22\]](#)；而威顿的证明所依赖的旋量结构在高维时空中并不普遍存在，因而也不能直接推广到高维情形。尽管困难重重，人们仍在这方面作着艰难努力。2006年，洛卡姆（Joachim Lohkamp）发表了一篇文章，试图将舍恩和丘成桐的证明推广到八维以上时空。可惜的是，在这一工作中他不得不重新引进了 $R \geq 0$ 这一阴魂不散的额外条件。这表明，即便他的证明本身无误，其结果也并不具有普遍性。

除上面提到的推广外，数学物理学家们还考虑了在某些特定的非渐近平直时空——比如渐近AdS时空（Anti-de Sitter spacetime）——中的

正质量定理。另外，除了舍恩和丘成桐以及威顿的经典证明外，也有人尝试用其他方法来证明正质量定理。比如彭罗斯、索金（Rafael Sorkin）和伍尔加（Eric Woolgar）在1993年提出过一种以时空的因果结构（具体地讲是以类光测地线的性质）为基础的新证明^[23]。这些我们就不再介绍了。

注释

^[1] 不过，人们并未停止这方面的尝试，前面提到的那些可能的做法每一种都仍有人在研究。也有人在研究其他引力理论中引力场的能量动量分布。

^[2] 其中最著名的事例来自脉冲双星PSRB1913+16，这也是人类最早发现的脉冲双星。1993年，它的发现者美国天体物理学家赫尔斯（Russell Hulse, 1950—）和泰勒（Joseph Taylor, 1941—）获得了诺贝尔物理学奖。

^[3] 除ADM质量外，在广义相对论研究中还会用到其他一些质量概念，比如邦迪（Bondi）质量。对于那些质量也存在相应的正质量问题或正质量定理。限于篇幅，本章将只讨论ADM质量。

^[4] 这里虽然用了“非物理”这一称谓，但事实上“非物理时空”的边界完全是由物理时空的共形性质所决定的，从这个意义上讲它还算“物理”。

^[5] 以度规张量为例，它在 i^0 只具有连续性以及一阶导数对方向的连续依赖性。

^[6] 需要提醒读者的是，虽然渐近平直时空的定义本身不存在歧义，但不同物理文献所采用的定义彼此间却往往是互有歧义的，即并不存在唯一的标准定义。不过那些定义之间的细微差异对于多数侧重物理的研究来说并无重大影响。

^[7] 了解外曲率的一个很好的例子是圆柱面，许多初学者在学习微分几何的时候，恐怕会对圆柱面的曲率张量为零感到困惑。因为在直觉上，圆柱面分明是“弯曲”的。这一直觉其实没有错，只不过它是来自于圆柱面所嵌入的三维空间，而非圆柱面本身，因而体现的是圆柱面的外曲率，而非内蕴曲率。倘若我们不是生活在三维空间，而是生活在圆柱面上，并且只能局地地感知圆柱面，那就不会觉察出它与平面的差异，从而也就不会产生诸如“圆柱面是弯曲的”那样的直觉。

[\[8\]](#) 有读者或许会问：出现在 K_{ij} 中的时间基矢的分解系数 N 和 N^i 是否也是动力学变量？答案是否定的，因为我们在3.3节中即将看到，广义相对论的动力学是有约束的动力学， N 和 N^i 是与约束条件相对应的拉格朗日（Lagrange）乘子。

[\[9\]](#) 人们通常将三元组 (Σ, h_{ij}, K_{ij}) 称为初始数据集（initial data set）。而如果 Σ 渐近平直，且在 i^0 满足一定的解析条件（这种解析条件对物质场的渐近行为也是一种约束），则相应的初始数据集被称为渐近平直的初始数据集。

[\[10\]](#) 这一能量（即ADM能量或ADM质量）虽然来自于对引力场作用量的分析，但它包含了物质场的贡献，因为后者已体现在了 h_{ij} 的渐近行为中（在弱场近似下，这一点可以通过直接计算予以证实）。ADM质量由渐近平直区域中的面积分表示，体现了一个朴素并且实用的物理思想，那就是确定一个孤立体系总能量的方法，是从远处观测其引力效应。另外，从这一表述可以看到，如果 Σ 是一个封闭曲面（比如 S^3 ），那么由于其边界为零， E 必定为零。人们有时把这一结果表述为：封闭宇宙的总能量为零。不过，这种表述的物理意义是值得怀疑的，因为在不存在渐近平直区域的情况下，总能量这一概念本身就很难有合适的定义。

[\[11\]](#) 早在正质量定理被证明之前，人们就已经知道，闵科夫斯基时空起码是准稳定的（quasi-stable），即它的能量相对于其他时空是（泛函意义下的）局部极小值。这是因为对闵科夫斯基时空的二阶微扰是所谓的引力波，它的能量早已被证明是正的。不过，如果正质量定理不成立，那么闵科夫斯基时空仍有可能是不稳定的，因为它有可能会通过非微扰过程衰变为ADM质量为负的时空。

[\[12\]](#) 这里我们假定牛顿引力只取决于物质，而不受引力场自身能量的影响，这是牛顿引力作为线性理论的重要特点。

[\[13\]](#) 我们在2.1节中曾经说过，奇点并不存在于物理时空中，因此这里所说的“经过奇点”实际上是指无限接近（但不包含）奇点。感兴趣的读者可以思考一下，渐近平直三维类空超曲面无限接近（但不包含）奇点 $r=0$ 为什么会破坏ADM质量的定义？

[\[14\]](#) 严格来讲，超引力理论中的正质量定理只是在形式上得到了证明，它的真正证明其实还根本不存在。这是因为超引力理论既不是一个经典理论，也不是一个可重整的量子理论，它的真正含义，以及它的包括能量动量在内的许多物理量的严格定义，直到今天也还是有待澄清的问题。因此只能说其证明“从表观上看”容易得多。

[\[15\]](#) 平均外曲率的定义有时与 $\text{tr}K$ 差一个常系数。

[\[16\]](#) 这里所述的步骤与舍恩和丘成桐的原始论文对证明步骤的分解略有差异，本文所述的第一步相当于舍恩和丘成桐原始论文中前两步的合并，而舍恩和丘成桐原始论文中的第三步则被我们分解成了两步。

[\[17\]](#) 需要强调的是，超引力理论及格里沙鲁的猜测对威顿的证明虽然有着启示作用，但威顿的证明并不依赖于超引力理论或格里沙鲁的猜测在现实世界中的适用性。也正因为如此，尽管超引力理论的ADM质量非负——如我们在3.4节中所说——只是一个形式上的结果，它能在经典极限下过渡为经典广义相对论中的正质量定理也只是一个猜测，但威顿对正质量定理的证明却是一个严格证明。

[\[18\]](#) 这里略去了对我们的介绍没有实质影响的常数因子，不同文献对这类因子的约定各不相同。

[\[19\]](#) 如果 $\varepsilon=0$ ，则可以证明不存在满足上述狄拉克方程的非平凡解。

[\[20\]](#) 相应地，满足这一额外条件的正质量定理也常被称为黎曼正质量定理。不过德国数学家黎曼（Bernhard Riemann，1826—1866）跟这些冠着他名字的结果并没有直接关系。

[\[21\]](#) 这里所说的时间（即1997年和1999年）是这两项工作实际完成的时间。它们直到2001年才正式发表在知名学术刊物《微分几何杂志》（Journal of Differential Geometry）上，其中布雷的文章发表在9月刊上，惠斯肯和伊尔玛能的文章发表在10月刊上。

[\[22\]](#) 2007年5月，布雷与李（Dan Lee）采用了类似于舍恩和丘成桐的技巧，将布雷本人所证明的黎曼彭罗斯猜想也推广到了八维及八维以下时空。

[\[23\]](#) 他们的这一工作响应者寥寥无几，除作者之一的伍尔加本人曾试图对这一证明进行推广外，极少有其他文献提及这一工作。

第4章 宇宙监督假设

4.1 黑洞、裸奇点及宇宙监督假设

在2.5节中，我们介绍了霍金-彭罗斯奇点定理。按照这一定理，奇点的形成在经典广义相对论中几乎是无可避免的。但奇点的存在对广义相对论的动力学有着很消极的影响。以最常见的施瓦西奇点为例，时空的曲率在奇点附近会趋于发散，物理学定律在那样的极端条件下将不再适用。即便对于性情比较“温和”——即时空曲率不发散——的奇点，它所具有的测地不完备性可以导致粒子在有限时间内从时空流形中消失，对于传统的物理学定律来说依然是一种破坏。虽然奇点本身——如我们在2.1节中所说——并不存在于物理时空之中，但如果由它造成的物理定律的破坏可以对物理时空中的演化产生影响，那么这种“借尸还魂”般的影响就足以使得我们无法有效地预言物理时空中的演化。

显然，奇点有可能具有的这种不良品性对广义相对论是一种威胁。因为对广义相对论来说，预言物理时空中的演化是一项重要使命。而从原则上破坏这种预言能力，则无疑是对广义相对论的一种巨大的，甚至堪称是颠覆性的破坏，这不仅是物理学家们不希望看到的，而且——在某些物理学家看来——也是上帝他老人家不希望看到的。

那么，在奇点本身几乎无可避免的情况下，有什么办法能够避免奇点有可能带来的破坏作用呢？这便是我们要在本节及接下来的几节中关注的问题。为了对这一问题及可能的解决方式有一个初步认识，我们先来看一个大家熟悉的例子：施瓦西奇点。我们知道，对于施瓦西奇点来说，大自然以一种非常有效的手段掩盖住了奇点所具有的不良品性，那就是：施瓦西奇点总是被包裹在所谓的施瓦西视界之内，而施瓦西视界

之内的区域（被称为施瓦西黑洞）与外部时空在因果上是完全隔绝的。这表明由施瓦西奇点所造成的任何物理定律的破坏都被严严实实地掩盖在了黑洞（或视界）之中，而不会影响外部时空中的演化。从这个意义上讲，施瓦西奇点的存在是无害的。

但施瓦西奇点只是最简单的奇点，它被包裹在黑洞（或视界）之内。这一性质能在多大程度上代表奇点的一般性质？特别是，当施瓦西奇点所具有的各种对称性不复存在的时候，奇点是否仍具有这种被黑洞（或视界）所包围的特性？这些却都是未知数。一个奇点倘若不被黑洞（或视界）所包围，它所造成的物理定律的破坏就有可能对物理时空中的演化产生影响，那样的奇点被称为裸奇点。为避免奇点的存在破坏时空中的演化性质，裸奇点的存在是必须被排除的。而排除裸奇点的一个重要途径，就是证明黑洞（或视界）的出现具有普遍性，而不依赖于任何特殊的对称性。

读者也许还记得，我们在1.1节中曾经介绍过有关奇点的出现是否依赖于对称性的争论。奇点定理的证明很漂亮地解决了那一争论。现在的问题几乎是当年那场争论的翻版，只不过当年所争论的是奇点的产生是否依赖于对称性，而我们现在所讨论的则是黑洞（或视界）的产生是否依赖于对称性。很多物理学家希望对后者也能找到一个像当年奇点定理那样的普遍答案，即凡能产生奇点的物理条件，也一定能产生包裹奇点的黑洞（或视界）。显然，这样的答案倘若存在，我们就不必担心奇点的出现会破坏时空中的演化性质了。这一良好愿望被称为宇宙监督假设，它是1969年由彭罗斯提出的。通俗地讲，宇宙监督假设要求所有的奇点都受到黑洞（或视界）的“监督”（即不存在裸奇点）。它的铁杆支持者霍金曾将这一假设幽默地表述为：上帝憎恶裸奇点（God abhors a naked singularity）。

可惜的是，与奇点的形成不同的是，对于黑洞（或视界）的形成，目前还没有普遍的结果。当然，这也不难理解，因为我们将会看到，黑

洞（或视界）的定义与时空的整个未来演化有着密切关系，因而不像奇点的定义那样局域，研究黑洞（或视界）的形成也因而要比研究奇点的形成困难得多。有些物理学家甚至不无悲观地认为，对那些性质的研究超出了人们迄今在微分几何及微分方程方面所具有的能力。这种悲观看法是否有道理，还有待观察，目前物理学家们所知道的是：描述黑洞的某些时空——比如施瓦西时空和克尔（Kerr）时空——在线性微扰下是稳定的。这是20世纪70—80年代由维希外希瓦拉（C. V. Vishveshwara）、普赖斯（Richard Price, 1943— ）、凯（Bernard Kay）、沃尔德（Robert Wald, 1947— ）和怀廷（Bernard Whiting）等人所证明的。这些结果意味着如果物质分布非常接近于形成施瓦西时空或克尔时空所需要的对称性，则它们的坍缩将会产生黑洞而非裸奇点。这一结果虽有很大的局限性，但对宇宙监督假设乃是很重要的早期支持。

另一方面，虽然裸奇点不受欢迎，但我们必须看到，在普遍意义下摒弃裸奇点是不现实的。事实上，按照大爆炸宇宙论，我们所生活的宇宙始于约138亿年前的一次大爆炸。从经典广义相对论的意义上讲，大爆炸正是一个裸奇点，我们不仅可以观测到它的残留效应，而且我们及我们周围的一切本身就是残留效应的一部分。不过大爆炸乃是既成事实，只要此后的时空演化不再产生新的裸奇点，广义相对论在实际意义上的预言能力就不会受到破坏。因此，在考虑宇宙监督假设时，像大爆炸这样的初始奇点需要被排除在外⁽¹⁾。宇宙监督假设关心的乃是在正常的物质性质及初始条件下，通过诸如引力坍缩之类的演化过程能否产生出裸奇点。

如果裸奇点不存在，那么所有的奇点就都应该被黑洞（或视界）所包围。因此在进一步讨论宇宙监督假设之前，让我们先对黑洞（或视界）的含义作一点讨论。在广义相对论中，黑洞这一概念是在渐近平直时空——或者更确切地说，是在所谓强渐近可预测（strongly

asymptotically predictable) 时空——中进行定义的⁽²⁾。什么叫做强渐近可预测时空呢？它是一种特殊的渐近平直时空⁽³⁾，在其中类光无穷远的因果过去——通常记为 $J^-(j^+)$ ，其中 J^- 表示因果过去， j^+ 表示类光无穷远——是全局双曲的⁽⁴⁾。由于“类光无穷远的因果过去”是由所有可以与类光无穷远建立因果联系的时空点所组成，而“全局双曲”——如我们在2.4节中所说——意味着时空中的演化可以通过适当的初始条件来预言。因此在强渐近可预测时空中只要给定适当的初始条件，我们就可以对所有能用光信号到达类光无穷远的部分——即 $J^-(j^+)$ ——做出完善的预言。

那么什么是黑洞呢？黑洞是由所有无法与类光无穷远建立因果联系的时空点所组成的时空区域（这正是“连光也无法从黑洞中逃逸”这一通俗说法的物理表述）。如果用 M 表示整个时空流形，那么所谓黑洞就是指 $M - J^-(j^+)$ （感兴趣的读者可以思考这样一个问题：我们把黑洞定义为由无法与类光无穷远建立因果联系的时空点组成的时空区域，这是否意味着黑洞可以与黑洞外有限远的时空点建立因果联系？）。黑洞的边界则被称为视界，或者确切地说是事件视界（event horizon），以区别于其他一些视界概念，比如2.4节中提到的柯西视界。显然，黑洞与视界的上述定义与时空的整个未来演化有着密切关系，因为对类光无穷远及能与之建立因果联系的点集的确定，都有赖于时空未来演化的整体性质。

将黑洞的概念与强渐近可预测时空的定义联系起来，我们可以看到，在强渐近可预测时空中，除黑洞以外的所有区域都具有良好的因果性质（全局双曲）。这样的区域显然不可能包含奇点。换句话说，在强渐近可预测时空中如果存在奇点，则奇点必定存在于黑洞之中，这正是宇宙监督假设所预期的性质⁽⁵⁾。因此，利用强渐近可预测时空这一概念，宇宙监督假设——确切地说是所谓的弱宇宙监督假设（weak cosmic censorship hypothesis），以区别于后文将会提到的强宇宙监督假

设（strong cosmic censorship hypothesis）——可以粗略地表示为：

宇宙监督假设（粗略版）：在正常的物质性质及初始条件下，时空是强渐近可预测的。

不过，这一表述虽然对宇宙监督假设的结论部分做了比较精确的表述（即时空是强渐近可预测的），却没有对前提部分，即所谓“正常的物质性质及初始条件”做出明确界定，因此充其量只是一个“半拉子工程”。那么，为了使宇宙监督假设成立，究竟什么才是所需要的“正常的物质性质及初始条件”呢？不幸的是，这是一个极其困难的问题——事实上，这个问题和宇宙监督假设本身一样，是一个迄今尚未完全解决的问题。但是，这个极其困难的问题对于宇宙监督假设的研究又是至关重要的，从而无法回避——因为若不假定“正常的”初始条件，就无法将初始奇点及其他一些奇巧的情形排除在外；而若不假定“正常的”物质性质，则可以轻而易举地用纯几何的方式构造出有裸奇点的时空，然后通过逆用爱因斯坦场方程来定义相应的物质分布（这样定义的物质分布通常具有奇特的性质），从而构造出宇宙监督假设的反例。从这个意义上讲，宇宙监督假设是一个双重难题，它的表述本身就是一个难题，它的证明则是难上加难。

值得一提的是，在探索宇宙监督假设的征途上，物理学家们除发挥智慧外，也充分展示了他们独特的幽默感。1991年9月24日，霍金与同事普雷斯基尔（John Preskill, 1953— ）及索恩以宇宙监督假设为对象打了一个赌（当然，霍金把赌注压在了自己强力支持的宇宙监督假设的成立上）。按照赌约，输家必须向赢家提供足以覆盖后者裸体的衣服（看来奇点是否受到监督虽还不得而知，赢家的裸体是必须受到衣服“监督”的），并在衣服上绣上愿赌服输的话。霍金这人赌运总体上讲是比较差的，但他的赌品倒是可圈可点。6年后，随着一些我们在后文中将会介绍的对宇宙监督假设不利的证据得到确立，霍金公开承认了自

己的落败。不过，在盘点了赌场失意的原因后霍金表示，自己之所以失利，乃是因为在赌约——即对宇宙监督假设的表述——中考虑不周，未对初始条件——即我们上文提到的“正常的初始条件”——作出足够的限定。换句话说，霍金认为自己是栽在了我们上文提到的“半拉子工程”上。当然，这番辩白在赢家普雷斯基尔及索恩看来纯属借口，于是这三位老顽童在霍金对赌约略作调整的基础上重开了赌局。这一年，即1997年，他们打赌的消息出现在了《纽约时报》（The New York Times）的科学专栏中，成为了大众新闻。

在接下来的几节中，我们将讨论一些与宇宙监督假设有关的论据，其中有些论据有利于宇宙监督假设，有些则不利。我们将通过对这些论据（尤其是不利论据）的讨论，来试图完善宇宙监督假设的表述。

4.2 摧毁黑洞——不可能的任务？

对宇宙监督假设的早期考察涉及一类很特殊的努力，那就是摧毁黑洞（或视界，下同）——当然只是通过理想实验。那种努力的基本思路是：既然宇宙监督假设要求所有奇点都被黑洞所包围，那么假如有办法在不破坏奇点的情况下将黑洞摧毁，原先遁迹其中的奇点岂不就暴露在了“光天化日”之下，从而变成裸奇点了？如果那样，宇宙监督假设就被推翻了。对于一个假设，能证明它成立固然很好，但推翻它显然也是一种解决方式。而且反过来说，即便无法推翻，只要努力“推”了，也算是积累了证据。因此，这种努力不失为考察宇宙监督假设的一个直观并且有益的研究方向。

那么，黑洞有可能被摧毁吗？

摧毁黑洞，这堪称是前所未闻的故事。事实上，在只有施瓦西解的情形下，这个问题或许连提都不会被提出。因为我们都知道，由施瓦西解所描述的施瓦西黑洞在经典广义相对论中是不可能被摧毁的。它就像一个贪婪而吝啬的守财奴，只认得一件东西，那就是质量。而且它对质量向来是只知索取，却绝不付出，它的视界当然也绝不会凭空消失，而只会像守财奴的钱包一样越来越大。但是，施瓦西黑洞只是最简单的黑洞，自施瓦西黑洞之后，物理学家们在广义相对论中又陆续发现了一些更复杂的黑洞解，它们不仅呈现出丰富多彩的性质，而且也为摧毁黑洞提供了一些可能的途径——当然，只是“可能”而已。

那么，那些更复杂的黑洞是什么样子的呢？研究表明，对于引力与电磁耦合的所谓爱因斯坦-麦克斯韦体系——也称为电真空

（electrovac）体系，广义相对论的所有稳定解（stationary solution）都可以由所谓的克尔-纽曼（Kerr-Newman）度规所描述^[6]。这一度规只带三个参数：质量 m 、电荷 Q ，以及单位质量所带的角动量 J 。这一结果

被称为黑洞无毛发定理（no hair theorem），或黑洞唯一性定理（uniqueness theorem）。对于我们的目的来说，克尔-纽曼度规的一个令人瞩目的特点是它有两个视界，分别位于径向坐标

$$r_{\pm} = m \pm (m^2 - Q^2 - J^2)^{1/2} \quad (4.2.1)$$

处。其中位于 $r=r_+$ 的被称为外视界，它是事件视界，这一视界以内的区域被称为克尔-纽曼黑洞；而位于 $r=r_-$ 的被称为内视界，它是柯西视界。假如电荷与角动量都为零，则内视界消失，外视界等同于施瓦西视界。除此之外，克尔-纽曼度规还有一个重要特点，那就是在赤道面上有一个等效半径为 J 的奇环。

由上述视界半径公式，细心的读者也许自己就能看出一个问题，那就是内外视界在 $Q^2 + J^2 = m^2$ 时将趋于重合，而在 $Q^2 + J^2 > m^2$ 时则会失去意义，因为视界公式中的被开方式将会变成负数。从物理上讲，这时候克尔-纽曼度规将不存在视界，从而也将不再描述黑洞。但克尔-纽曼度规所具有的奇环却依然存在，这个奇环此时就变成了所谓的裸奇环，它的出现将破坏宇宙监督假设。这对于我们雄心勃勃的摧毁黑洞计划来说，无疑是一条重要线索。看来贪婪的黑洞似乎也有弱点，这弱点正是贪婪！通俗地讲，我们只要设法让一个它吃进过量的电荷或角动量，使得 $Q^2 + J^2 > m^2$ ，就能把它“撑死”。

这计谋看起来不错，但关键是：有可能得逞吗？这种源源不断地向黑洞输送电荷或角动量，直至将之摧毁的过程真的有可能实现吗？关于这一点，物理学家们曾经作过分析，但结果——很遗憾地——却是否定的。否定的理由其实很直观，下面介绍一下。为简单起见，我们只考虑通过获得过量的电荷来摧毁一个不旋转黑洞——即所谓的雷斯勒-诺斯特朗姆（Reissner-Nordström）黑洞——的努力。假定我们每次向黑洞投放电荷 δQ ，携带这一电荷的粒子质量则为 δm 。为了让黑洞的电荷增加快于质量增加，我们要求 $\delta Q > \delta m$ 。向黑洞投放这样的电荷在一开始是

很容易的，事实上，我们只要将电荷放在黑洞周围，它就会自动地被黑洞的引力所俘获。可惜这样的好光景并不能持久，随着黑洞的总电量 Q 越来越接近总质量 m ，继续向黑洞投放电荷就会变得越来越困难，因为黑洞中已有的电荷将会对新投放的电荷 δQ 产生越来越强烈的排斥，这种排斥最终将会超过质量 δm 所受的引力（请读者想一想为什么）。在这种情况下，为了让新投放的电荷能被黑洞俘获，我们就不能简单地将电荷放在黑洞周围，而必须使劲地将它扔向黑洞，让它依靠初速度来克服来自黑洞电荷的排斥作用。但是这样做的副作用却是增加了携带电荷的粒子所具有的初始能量，即增加了 δm 。1974年，沃尔德计算了这一初始能量的大小，结果发现若所讨论的黑洞已经处于 $Q=m$ 的极端带电状态，那么向这一黑洞投放电荷 δQ 所提供的初始能量 δm 必须满足

$$\delta m \geq \delta Q \quad (4.2.2)$$

由此可见，为了能在这种极端情形下继续向黑洞输送电荷，被输送的电荷所具有的能量将会自动保证 $m \geq Q$ 继续得到满足。换句话说，黑洞是不会因为上面这种输送电荷的方法而被摧毁的。

类似地，如果我们试图通过向一个 $J=m$ 的极端旋转黑洞投送具有很大轨道角动量的粒子以突破 $m \geq J$ ，则该粒子所具有的碰撞参数（impact parameter）将大到使之无法击中黑洞。而如果我们试图投送的是一个具有很大自转角动量的粒子，则该粒子与旋转黑洞之间会产生自旋-自旋相互作用，这种相互作用同样会阻止其击中黑洞。将上面这些分析综合起来，我们看到，通过向黑洞投放电荷或角动量来摧毁黑洞乃是不可能任务（mission impossible）[\(7\)](#)。大自然在极端黑洞情形下显示出的保护黑洞的手段是令人瞩目的，简直像是在有预谋地阻止裸奇点的出现，这是支持宇宙监督假设的一类很直接的证据[\(8\)](#)。

4.3 彭罗斯猜想与宇宙监督假设

对破坏宇宙监督假设的可能性的另一类考察是彭罗斯所做的。他考虑了一个渐近平直超曲面 S 上满足主能量条件的某种物质分布^[9]。视具体情况而定，这种物质分布有可能产生封闭陷获面（封闭陷获面的定义请参阅2.4节）。如果宇宙监督假设成立，那么可以证明，所有这些封闭陷获面都必须被包含在黑洞之内，我们用 A 表示黑洞的外视界面积。在物理上可以预期，这一黑洞在经历了各种经典广义相对论所允许的变化——比如吸积物质，辐射引力波，等等——之后，最终将会进入某种稳定状态，即变成一个稳定的黑洞。按照黑洞无毛发定理，这一稳定的黑洞必定是克尔-纽曼黑洞。假定这一克尔-纽曼黑洞的外视界面积为 A_1 。我们知道，霍金在1971年曾证明过一个著名的定理，叫做黑洞面积定理（blackhole area theorem），它表明如果宇宙监督假设成立，且物质分布满足零能量条件，那么黑洞的视界面积永不减小^[10]。按照这一定理，显然有 $A \leq A_1$ 。

另一方面，克尔-纽曼黑洞的外视界面积为 $A_1 = 4\pi (r_+^2 + J^2)$ （其中 r_+ 为外视界的径向坐标值，由式（4.2.1）给出）。不难证明（请读者自行验证），克尔-纽曼黑洞的质量与视界面积之间满足： $A_1 \leq 16\pi m^2$ 。由于 $A_1 \leq 16\pi m^2$ 正是质量 m 的施瓦西黑洞的视界面积，因此这个结果的物理意义很简单，那就是在具有相同质量的所有黑洞中，施瓦西黑洞具有最大的视界面积。将这一结果与上面的不等式 $A \leq A_1$ 联立起来，我们就得到 $A \leq 16\pi m^2$ 。

如果我们进一步假设在整个过程中没有外来的质量（由于渐近平直超曲面 S 实际上就是全空间，因此这一假设显然是合理的），那么克尔-纽曼黑洞的质量显然不可能大于 S 上的ADM质量 M ，即 $m \leq M$ （ADM质量的定义请参阅3.3节）。将上述所有环节联系起来，我们可以看到，

如果宇宙监督假设成立，那么S上的初始条件必须满足一个不等式： $A \leq 16\pi M^2$ ，其中A为S上所有黑洞的外视界面积之和，M为S上的ADM质量。

读者们想必认出来了，这个不等式正是我们在3.7节中介绍过的彭罗斯猜想。1973年，彭罗斯正是通过类似于我们这里所介绍的思路而提出这一猜想的。从上述思路中可以看到，彭罗斯猜想几乎可以视为宇宙监督假设的推论，或者说几乎可以视为宇宙监督假设成立的必要条件（但没有任何迹象表明它有可能是充分条件）。在4.1节中我们已经看到，宇宙监督假设远比奇点定理来得困难，而通过上面的介绍，我们又可以看到，宇宙监督假设要比彭罗斯猜想更为困难。当然，上述介绍并不构成对两者关系的数学证明，因为我们用到了一些未予严格证明的假定，比如假定所有黑洞最终都会变成克尔-纽曼黑洞。但即便如此，它仍然为寻找破坏宇宙监督假设的努力提供了一个新的可能方向，那就是寻找能破坏彭罗斯猜想的初始条件。

不过，从我们在3.7节所介绍的情况来看，虽然彭罗斯猜想目前还只是一个猜想，但人们在研究这一猜想上已经取得的进展表明，它最终被证明的可能性是很大的。相应地，通过寻找能破坏彭罗斯猜想的初始条件来构筑宇宙监督假设的反例，其希望则是很渺茫的。对于相信宇宙监督假设的人——比如霍金——来说，这显然又是一条好消息，因为彭罗斯猜想的成立虽然并不意味着宇宙监督假设一定成立，但无疑构成一种很强的支持——起码，它可以排除一大类破坏宇宙监督假设的可能性。也正因为有这么多的好消息，霍金才会信心满满地下注。只可惜，好消息虽多，却无法构成证明，而坏消息哪怕只有一条，也足以带来麻烦。

4.4 壳层穿越奇点与壳层会聚奇点

由于黑洞无毛发定理的存在，以及摧毁黑洞的努力所遭遇的困难，利用广义相对论的稳定解来寻找裸奇点的可能性起码在经典广义相对论的范围内可以说是基本被排除了⁽¹¹⁾。不过，对裸奇点的直接寻找依然是研究宇宙监督假设的主要途径之一，只是进一步努力的方向在很大程度上转向了动态解。这其中，动态但具有良好对称性的解要比缺乏对称性的解容易研究得多，从而成为了最早获得突破的方向。

在广义相对论的所有具有良好对称性的动态解中，最简单的无疑是球对称动态解。不过我们知道，球对称引力场本身是不具有动力学自由度的——这是所谓的伯克霍夫定理（Birkhoff theorem），它表明真空中的球对称度规必定是施瓦西度规⁽¹²⁾。因此，非平凡的球对称动态解必须涉及物质，从而必须涉及由物质能量动量张量带来的额外复杂性。作为起步，最明智的做法显然是选择尽可能简单的能量动量张量。那么，什么样的物质具有最简单的能量动量张量呢？答案是理想尘埃物质，这种物质的能量动量张量为 $T^{\mu\nu} = \rho u^\mu u^\nu$ ，它是理想流体能量动量张量 $T^{\mu\nu} = p g^{\mu\nu} + (p + \rho) u^\mu u^\nu$ 在压强恒为零时的特例（由于这一缘故，理想尘埃也被称为零压理想流体），只带一个与物性有关的标量函数 ρ 。因此研究理想尘埃物质的球对称动态解成为了寻找裸奇点的最佳切入点之一。

另一方面，在所有动态过程中，与奇点的形成最有关系的显然是坍缩过程，因此在所有球对称动态解中，描述球对称坍缩过程的解是最值得关注的。当然，球对称坍缩过程倘若在引力的支配下进行到底，其最终产物将是施瓦西黑洞，从而是不带裸奇点的。但在坍缩的过程中，即尚处于动态的阶段中，却完全有可能出现复杂的情况。特别是，如果我们能找到一个球对称坍缩过程，使得当奇点在某个半径——包括坍缩中

心——出现时，包裹该奇点的任一同心球面之内都尚未积累起使该球面成为施瓦西视界所需的物质（从而施瓦西视界尚无法形成），则这种先于视界而出现的奇点就是裸奇点——虽然它只在有限的时间内才是裸露的⁽¹³⁾。

这样的裸奇点究竟会不会出现呢？凭空想象是得不出结果的，答案只能来自计算。这方面最早的计算比宇宙监督假设早了整整30年就已经出现了。1939年，美国物理学家奥本海默（Robert Oppenheimer, 1904—1967）和他的学生施奈德（Hartland Snyder, 1913—1962）研究了最简单的球对称坍缩过程：一个均匀理想尘埃球的坍缩。他们发现，这样的坍缩在随尘埃物质一同运动的观测者——即所谓的随动观测者

（comoving observer）——看来，将会在有限的时间内完成，这其中奇点只有在坍缩过程彻底完成时才会出现，而这时的时空已经完全变成了施瓦西时空。不仅如此，其实早在奇点出现之前，一旦尘埃物质进入到相应的施瓦西半径以内，视界就已经形成。因此，均匀理想尘埃球的坍缩是一个视界先于奇点而形成的例子，它显然不会产生裸奇点。

看来，这又是一个有关宇宙监督假设的“利好消息”，但这一次它同时也是“利空消息”的邻居。

奥本海默与施奈德的计算如今已是每位学习广义相对论的学生都能完成的“小习题”。但广义相对论计算的困难之处就在于，哪怕是在这样的“小习题”当中加入一些额外因素，比如非均匀、有压强（即不再是尘埃物质）或有自转（即不再是球对称），问题就会急剧复杂化，甚至复杂到让奥本海默那样的人物也不得不退避三舍。这方面的研究直到30多年后才有了新的突破。1973年，德国汉堡大学的物理学家约德杰斯

（Peter Yodzis）、塞弗特（Hans Seifert）等人再次对理想尘埃球的坍缩进行了研究。在他们的研究中，压强仍保持为零，物质的分布也仍维持球对称，唯一被松绑的只有均匀性条件。可以说，他们只是把潘多拉盒子打开了一条小小的缝隙。

从这条小小的缝隙里，他们会发现什么呢？

约德杰斯等人发现，当尘埃物质的密度分布满足一定的条件时⁽¹⁴⁾，不同壳层上尘埃物质的坍缩速度会出现显著差异——特别是，快速坍缩的外壳层可以穿越缓慢坍缩的内壳层。当两层尘埃物质彼此穿越时，某些物理量或几何量会出现奇异性，而此时物质的整体坍缩尚未进行到足以形成视界的程度。换句话说，在这种特殊的坍缩过程中，奇点的出现可以早于视界的形成，这意味着裸奇点出现在了坍缩过程中！约德杰斯等人的这一结果是人们最早发现的能导致裸奇点的坍缩过程之一，这种由壳层穿越而导致的奇点（实际上是球对称的奇异面）被称为壳层穿越奇点（shell-crossing singularity）。研究表明，壳层穿越奇点是一种奇异性较弱的奇点。不过，宇宙监督假设并未对奇点奇异性的强弱进行限定（我们也因此而未对奇异性的强弱作具体介绍），因此约德杰斯等人的这一结果对于宇宙监督假设具有一定的冲击性。

不过，约德杰斯等人所采用的理想尘埃模型无可避免地削弱了这一结果的现实意义。因为理想尘埃的压强恒为零，而物理直觉告诉我们，在足以形成奇点的坍缩过程中，物质将会受到高度挤压，在这种过程中压强是绝不可能为零的。有鉴于此，约德杰斯等人很快就展开并完成了一系列后续研究。在那些研究中，他们将理想尘埃换成了理想流体（从而压强不再为零），而结果则不仅维持了裸奇点的存在，还进一步证明了裸奇点在球对称微扰下具有稳定性。

除约德杰斯等人所发现的壳层穿越奇点外，人们在球对称坍缩过程中还发现了一类比壳层穿越奇点更为棘手的裸奇点：壳层会聚奇点（shell-focusing singularity）。这类裸奇点位于坍缩球体的中心⁽¹⁵⁾，只要物质分布满足特定的物态及初始条件，它既可以出现在理想流体（包括理想尘埃）的坍缩过程中，也可以出现在所谓辐射物质的坍缩过程中，后者的能量动量张量为 $T^{\mu\nu} = \rho k^\mu k^\nu$ （其中 k^μ 为类光矢量）。研究表明，壳层会聚奇点的奇异性要高于壳层穿越奇点。

这样看来，在特定的物质坍缩过程中裸奇点的出现已是“铁证如山”的事实，宇宙监督假设尽管寄托了一些物理学家的良好愿望，并且也得到了不少旁证，却早在几十年前就该寿终正寝了。但事实上，这一假设却迄今仍被视为是一个尚未解决的问题，这是怎么回事呢？为什么如此直接的反例却未被认为是推翻了宇宙监督假设呢？这其中的诀窍还得到我们在4.1节中提到的那个“半拉子工程”的表述中去寻找。

4.5 走向严密表述

现在让我们回顾一下4.1节中提到的那个“半拉子工程”的表述：“在正常的物质性质及初始条件下，时空是强渐近可预测的。”我们当时就已指出，这个表述没有对“正常的物质性质及初始条件”做出明确界定，因而充其量只是一个“半拉子工程”。解决前述问题的诀窍就在这所谓“正常的物质性质”上——这个含糊其辞的条件简直就是为对付上面提到的那些“反例”而量身定做的。

在4.4节提到的两种裸奇点中，壳层穿越奇点相对来说是不足为患的，因为这种因内外壳层彼此穿越而产生的奇点在闵科夫斯基时空，甚至在非相对论流体力学中就已经存在，从而并不是广义相对论特有的。我们知道，在广义相对论之前的经典物理学中并不存在类似于奇点定理那样的结果，那里出现的奇点即便在经典物理本身的范围之内，也只不过是来自于对现实世界的过度理想化，而不具有基础意义。壳层穿越奇点既然在普通流体力学中就存在，它的起源也就昭然若揭了，那便是来自于理想流体对现实流体的过度理想化。现实流体虽然也有壳层穿越现象，但在穿越过程中是绝不会出现奇异性的。因此，出现在理想流体中的壳层穿越奇点只是理想流体这一物质模型的缺陷，而不是广义相对论的麻烦。这种本身就有问题，却想把糊涂账转嫁到广义相对论头上的物质模型，理所当然地遭到了广大爱好宇宙监督假设之士的“痛斥”，被定性为不具有“正常的物质性质”，而遭驱逐出境。

与壳层穿越奇点不同——从而更为棘手——的是，壳层会聚奇点并不总是出现在普通流体力学中，因此它起码可以部分地归因于广义相对论。但是，既然理想流体本身已经因为不具有“正常的物质性质”而遭驱逐，它所造成的一切麻烦——包括壳层会聚奇点——当然也就得一并卷铺盖了，正所谓“皮之不存，毛将焉附”，壳层会聚奇点傍错了老板，只

能怪自己命苦了。

细心的读者也许会问：辐射物质的坍缩不也可以产生壳层会聚奇点吗？难道连辐射也不具有“正常的物质性质”吗？答案是肯定的，不过不是因为辐射不具有正常的物质性质，而是因为在4.4节中被称为“辐射”的东西乃是“披着羊皮的狼”，是一种冒牌的辐射。它之所以被称为“辐射”，只不过是因其能量动量张量 $T^{\mu\nu} = \rho k^\mu k^\nu$ 中的 k^μ 类光。将这种“辐射”与理想尘埃的能量动量张量相对比，可以看到它所描述的物质相当于类光的“尘埃”，这是使用“辐射”一词的核心依据。由于类光矢量也叫零矢量，因此这种“辐射”也称为零尘埃（null dust）。可惜的是，“零尘埃”虽与普通辐射一样具有类光性，其能量动量张量却并不描述任何现实世界中已知的辐射，比如电磁波（要从数学上证明这一点虽不算困难，却也不是完全轻而易举的，不过请读者想一想，有什么简单的特性可以让我们从物理上判定像零尘埃这样的“辐射”并不是像电磁波那样的真实辐射？）。因此，这种挂着羊头卖狗肉的“辐射”并不具有“正常的物质性质”。

这样一来，4.4节提到的那些裸奇点就全都被扫地出门了，“正常的物质性质”这一“无招胜有招”的含糊术语算是在关键时刻挽救了宇宙监督假设。但是，这一仗却胜得并不漂亮，因为这种一味依靠含糊术语来打游击的做法终究是不够堂正，甚至可以说是有些无赖的。我们不禁要问：说了半天，究竟什么才是“正常的物质性质”？能不能对“正常的物质性质”给出一个正面的定义，而不要总等到反例逼上门来，才事后诸葛般地从反面告诉我们什么不是“正常的物质性质”？

这是一个合情合理的问题，也是一个合情合理的要求。幸运的是，物理学家们为这份合情合理找到了一个合情合理的答案。这个答案是这样的：如果用 $\varphi_1, \dots, \varphi_n$ 表示物质场或其分量，那么具有“正常的物质性质”的物质场被定义为满足以下形式的准线性二阶双曲运动方程组：

$$g^{\mu\nu}(x, \varphi_j, \varphi_{j;\rho}) \varphi_{i;\mu\nu} = F_i(x, \varphi_j, \varphi_{j;\rho}) \quad (4.5.1)$$

这里，准线性（quasilinear）指的是相对于最高阶（对于本方程组来说即二阶）协变导数为线性。这一定义参照了包括电磁场在内的物理学中所有基本动力学体系的主要特征，并且做了适当的推广（比如允许度规张量 $g^{\mu\nu}$ 依赖于物质场的一阶协变导数），因而被认为具有极大的普适性^[16]。更重要的是，早在1952年人们就已证明，满足式（4.5.1）的物质场在全局双曲背景时空中具有良好的初值表述（initial value formulation），即给定光滑类空柯西面上的一组初始条件（在本书中，若无特殊说明，凡初始条件都是指非奇异的），运动方程在该柯西面的一个邻域内存在唯一且连续依赖于初始条件的解。物质场所满足的这一性质保证了它在闵科夫斯基时空（那是全局双曲背景时空的最简单特例）中具有良好的动力学性质。而既然物质场在闵科夫斯基时空中具有良好性质，那么假如它与引力的耦合体系出现了问题（比如出现裸奇点），则那种问题就可以被认为是与引力的存在密切相关，而不是像壳层穿越奇点那样的“诬告”。这一特点使得上述定义很适合于宇宙监督假设的讨论。当然，对于宇宙监督假设来说，我们还要求物质场满足适当的能量条件，比如主能量条件，以排除诸如 $m < 0$ 的施瓦西时空那样的情形（ $m < 0$ 的施瓦西时空——如我们在3.4节中所说——是带有裸奇点的）。

现在我们已经接近对宇宙监督假设给出一个比较严格的表述了，不过为了完成这一表述，除“正常的物质性质”外，还必须对粗略定义中“正常的初始条件”也给出一个直接定义。这一点初看起来是不复杂的，因为考虑到黑洞和视界的定义有赖于时空的渐近平直性（参阅4.1节），所谓“正常的初始条件”，显然就应该定义为某个渐近平直类空超曲面 Σ 上的满足一定渐近条件的 h_{ij} 和 K_{ij} ——简单地讲，就是渐近平直的初始数据集。有关这一点，我们在讨论广义相对论的动力学时已经表述

过了，读者们可参阅3.2节，尤其是该节的注释。不过我们稍后将会看到，这“正常的初始条件”其实也是颇有玄机的。

综合上面的讨论，现在我们正式给出宇宙监督假设的一个比较严格的表述：

宇宙监督假设：在广义相对论中，如果

- (1) 物质场的运动方程组为形如式(4.5.1)的准线性二阶双曲方程组；
- (2) 物质的分布满足主能量条件；
- (3) (Σ, h_{ij}, K_{ij}) 为渐近平直的初始数据集。

则 (Σ, h_{ij}, K_{ij}) 的最大柯西展开是强渐近可预测的。

这里 (Σ, h_{ij}, K_{ij}) 的“最大柯西展开” (maximal Cauchy development) 指的是该初始数据集所对应的解 $(M, g_{\mu\nu})$ 在微分同胚 (diffeomorphism) 意义上保持 Σ 为柯西面的最大延拓⁽¹⁷⁾。

好了，现在我们终于有了一个关于宇宙监督假设的明确表述。宇宙监督假设也已经闯过了上面介绍过的所有关卡。那么，在现在这样一个明确表述下，它还能继续过关斩将吗？

4.6 零质量标量场与裸奇点

宇宙监督假设的明确表述虽然帮助宇宙监督假设闯过了若干关卡，却很不幸地排除了诸如理想流体之类广义相对论研究者们喜闻乐见的物质分布。为了进一步研究宇宙监督假设，物理学家们不得不引进其他类型的物质。这其中最简单的莫过于标量场（scalar field），也叫做克莱因-戈登场（Klein-Gordon field）。而标量场中最简单的则是所谓的零质量标量场，它的能量动量张量为 $T_{\mu\nu} = \varphi_{;\mu}\varphi_{;\nu} - (1/2)g_{\mu\nu}\varphi^{;\rho}\varphi_{;\rho}$ ，相应的运动方程则为最简单的线性二阶双曲方程（请读者自行证明）：

$$g^{\mu\nu}\varphi_{;\mu\nu} = 0 \quad (4.6.1)$$

这种标量场与引力耦合的体系被称为爱因斯坦-克莱因-戈登体系^[18]，它显然满足宇宙监督假设的明确表述对运动方程的要求，因而成为了研究宇宙监督假设的新战场。

自20世纪80年代中期开始，当时任教于美国雪城大学（Syracuse University）的数学物理学家克里斯多洛（Demetrios Christodoulou, 1951—）开始对爱因斯坦-克莱因-戈登体系进行较为系统的研究。由于问题的高度复杂性，克里斯多洛不得不将注意力集中到具有球对称性的爱因斯坦-克莱因-戈登体系上。但即便引进了这么高的对称性，问题依然相当困难，以至于与宇宙监督假设有关的突破直到6年之后的1992年才由当时在美国得克萨斯大学奥斯汀分校（University of Texas at Austin）的物理学家肖普推克（Matthew Choptuik）做出。肖普推克在克里斯多洛的一些前期解析研究的基础上，对球对称爱因斯坦-克莱因-戈登体系的坍缩过程进行了数值研究。他的研究表明，如果向内坍缩的标量场的某个强度参数足够小，坍缩场将会被反射回无穷远；而如果该强度参数足够大，坍缩场则会形成黑洞。这两种极端情形都不破坏宇宙监

督假设。但是，这两种极端情形之间还存在一段参数范围，那里的情况如何呢？特别是，当强度参数由大变小，在由能够形成黑洞转变为无法形成黑洞的那一个被称为临界参数（critical parameter）的特殊数值附近的情况又如何呢？这些问题克里斯多洛曾在1987年提出过，肖普推克找到了答案，他的结果发表于1993年初。

肖普推克的数值研究表明，当强度参数由大变小接近临界参数时，黑洞的质量 M_{BH} 与参数 p 之间存在一个幂函数关系： $M_{\text{BH}} \sim (p - p_c)^\gamma$ ，这里 p_c 为临界参数， γ 是一个对标量场具有普适性的指数，数值约为0.37。肖普推克的这个结果表明，在接近临界参数时，黑洞的质量——因而其视界——将无限缩小，外部观测者可以无限接近黑洞中心的奇点。这种可以无限接近的奇点已经非常类似于裸奇点，但还不完全是，因为对于任何给定的强度参数，只要它不完全等于临界参数，这种接近程度就终究还是有限的。那么，当强度参数完全等于临界参数时，情况又如何呢？这时候黑洞的视界将会消失，但奇点是否也会随之消失呢？如果奇点也消失了，那就仍然不存在裸奇点。幸运（或不幸——取决于看问题的立场）的是，肖普推克的研究表明，在这种临界情形下，时空的曲率标量在坍缩中心附近趋于发散，即奇点依然存在！因此，临界情形下爱因斯坦-克莱因-戈登体系的坍缩会在坍缩中心形成裸奇点。这是物理学家们首次在具有“正常的物质性质”的引力耦合体系中找到裸奇点。

比肖普推克的稍晚，印度塔塔基础研究所（Tata Institute of Fundamental Research）的物理学家德维外迪（Indresh Dwivedi）与乔希（Pankaj Joshi）给出了一个稍具一般性的结果，即在满足弱能量条件的球对称体系中，特定的非奇异初始条件可以导致裸奇点。而克里斯多洛本人也在爱因斯坦-克莱因-戈登体系的特定球对称坍缩过程中发现了性质与肖普推克所发现的临界情形相类似的裸奇点。1995年，剑桥大学的数学物理学家哈梅迪（Rufus Hamadé）与斯图尔特（John Stewart）通过

更细致的数值计算进一步证实了肖普推克的结果。

出现在零质量标量场中的这一系列结果对宇宙监督假设是沉重的打击（事实上它直接否证了我们在4.5节中所给出的宇宙监督假设的“比较严格”的表述）。虽然原则上我们仍可以有“遁词”，比如现实世界中并不存在具有基础意义的零质量标量场（粒子物理标准模型中唯一的标量场——希格斯场——是有质量的），但由于零质量标量场具备基本物质场所应具备的所有良好的动力学特征，因此我们有理由认为（虽然那不等同于证明），裸奇点在零质量标量场中的出现意味着“正常的物质性质”并不能成为挽救宇宙监督假设的救命稻草。

在这一系列的不利结果面前，久经考验的宇宙监督假设的捍卫者霍金终于扛不住了。1997年初，他——如我们在4.1节末尾所述——终于公开承认自己在与同事普雷斯基尔及索恩的赌局中落败，并依约向后者提供了赌注之一的足以覆盖对方裸体的衣服。不过，一局之败显然并不能让霍金就此俯首称臣。心有不甘的他请人在输给对方的衣服上绣上了那句著名的短语：“上帝憎恶裸奇点”，以示最终的胜负还得“走着瞧”。而且——也如我们在4.1节末尾所述——就在认输的同一天，他用一份修改了的赌约与普雷斯基尔及索恩开始了一个新的赌局。

霍金之所以要对赌约进行修改，除了因为对宇宙监督假设痴心不改外，还有一个很重要的原因，那就是他注意到人们当时针对宇宙监督假设所构造的反例全都依赖于非常特殊的初始条件，比如肖普推克的裸奇点要求标量场的强度参数恰好等于临界值，大了或是小了都不行；此外它们还大都依赖于非常严格的对称性，比如球对称性。如果宇宙监督假设的所有反例全都具有类似性质，即依赖于非常特殊的初始条件或严格的对称性，那它们在现实世界中实际上是无法出现的。有鉴于此，霍金在新赌约中规定：初始条件必须是一般的（generic）而不是特殊的。换句话说，霍金认为：一般初始条件（**generic initial condition**）下的动力学演化不会产生裸奇点^[19]。或者反过来说：裸奇点只能由特殊的

初始条件所产生。

显然，霍金的这一附加条件可以被认为是对4.1节的粗略表述中提及过，并在上一节的“比较严格”的表述中以渐近平直初始数据集的方式有所体现——但显然体现得还不够——的“正常的初始条件”的补充。从4.5节中我们知道，“正常的物质性质”曾一度挽救过宇宙监督假设，可惜挽救得并不彻底。现在霍金把新的希望寄托在了“正常的初始条件”上，希望它能干得比“正常的物质性质”更好，甚至一劳永逸地完成“救驾”任务。由于现实世界每时每刻、每个角落都充满着各种各样的随机微扰（比如宇宙微波背景辐射、引力波等），因此霍金寄希望于从初始条件的一般性这一角度来排除裸奇点，的确是一个比较合理的想法。新赌约对赌注的规定没什么大的改变，大家虽然都不缺钱，但衣服总是需要的，因此输家仍须向赢家提供足以覆盖后者裸体的衣服。不过普雷斯基尔与索恩对霍金上次所绣的“上帝憎恶裸奇点”看来有些耿耿于怀，因此在新赌约中特意强调衣服上所绣的必须是一句“真正愿赌服输”（truly concessionary）的话，而不能再玩花招。

4.7 讨论

霍金等人的新赌约问世后不久，1999年，上文提到的克里斯多洛在普林斯顿大学及高等研究院（Institute for Advanced Study）主办的《数学年报》（Annals of Mathematics）上发表了一篇文章（这篇文章的基本结果其实早在霍金等人的新赌约问世之前就已完成，并曾被圈内人士广泛引用）。在这篇文章中，他对爱因斯坦-克莱因-戈登体系做了一些解析研究，目的是想判定形成裸奇点的初始条件是否稳定。他的研究表明，对于爱因斯坦-克莱因-戈登体系来说，即便在球对称的情况下，能形成裸奇点的初始条件也是不稳定的，它会被特定类型的球对称微扰所破坏。显然，克里斯多洛的这一结果对霍金的新赌约是一个很大的支持，尽管它距离证明“霍金版”的宇宙监督假设无疑还差得很远。

霍金等人的新赌局迄今尚未见出输赢。这份赌约虽不是学术作品，却在很大程度上指出了宇宙监督假设的一个重要的研究方向。因为事实上，种种反例的存在已经推翻了如我们在4.6节中所表述的那种宇宙监督假设。要想挽救宇宙监督假设，附加条件的引进已是势在必行，而霍金在新赌约中引进的初始条件的一般性，在所有有可能挽救宇宙监督假设的附加条件中，几乎是唯一一种不至于将之弱化到失去现实意义的条件。不过，尽管霍金引进的附加条件不至于将宇宙监督假设弱化到失去现实意义，但却仍有可能造成一些不容忽视的后果。比如我们知道，广义相对论中有一些重要的结果——比如我们在4.3节中提到过的黑洞面积定理——是在类似于4.6节所给出的那种如今看来并不成立的宇宙监督假设的基础上证明的。霍金引进的附加条件对那些证明及其结论会产生什么影响？这也许是值得有志于广义相对论研究的读者去思考的。

在结束本专题之前，让我们回过头来提一下4.1节留下的一个伏笔。读者们也许还记得，我们在给出宇宙监督假设的粗略版时曾经提

到，那是所谓的弱宇宙监督假设。什么是弱宇宙监督假设？它是指只要求奇点存在于视界以内，从而与视界外的时空断绝因果联系，却并不排除视界内的观测者直接观测到奇点的可能性。但是，假如我们认为经典广义相对论必须自洽到不允许裸奇点具有观测意义的程度（这是人们提出宇宙监督假设的一个重要动机），那么视界内的观测者似乎也不应该例外。有人可能会说：视界内的观测者既然再也不可能回到视界之外了，那无论他们观测到什么，都已不再具有实际意义。这话初听起来不无道理，但我们别忘了，黑洞原则上是可以任意大的，可以大到不仅让观测者能平安穿越视界（黑洞越大，观测者穿越视界时所经受的潮汐力就越小），而且还可以让他们在黑洞内存在任意长的时间，甚至可以在那里从事系统的物理研究。在那里难道他们会发现完全不同的物理学吗？对于这样假想性的问题，不同的人显然会有不同的看法。但有一些物理学家相信，既然我们是由于对经典广义相对论的自洽性具有信心而提出了宇宙监督假设，就没有理由把视界内的观测者排除在外。他们提出，裸奇点不仅对于视界外的观测者，而且对于视界内的观测者都是不具有观测意义的。这就是所谓的强宇宙监督假设。

依据强宇宙监督假设，只要排除像大爆炸那样的初始奇点，时空的所有区域都应当能被适当的初始条件所预测，因此强宇宙监督假设曾被彭罗斯表述为：除去初始奇点外，所有物理上合理的时空都是全局双曲的。当然，对强宇宙监督假设的严格表述将会涉及一些数学上的微妙细节，以排除通过简单拓扑剪拼手段构造出的反例，这些我们就不在这里介绍了。

与弱宇宙监督假设的初看起来正确不同，强宇宙监督假设似乎初看起来就不太正确。我们早已知道，奇点在经典广义相对论中是无可避免的。由于强宇宙监督要求奇点对所有观测者都不具有观测意义，这意味着仅仅像弱宇宙监督假设所要求的那样用视界包裹奇点是不够的，为了使视界内的观测者也无法观测到奇点，奇点必须是类空的。类空奇点

的一个例子是施瓦西奇点，观测者虽然可以依据广义相对论预知其存在，但只有在真正撞上时才能“观测”到这样的奇点。但是，并非所有奇点都是类空的，比如由雷斯勒-诺斯特朗姆或更一般的克尔-纽曼度规描述的奇点就是类时的。对于那样的奇点，所有进入内视界的观测者都有可能直接观测到它的存在及可能的物理效应。

读者也许会觉得奇怪，既然像雷斯勒-诺斯特朗姆和克尔-纽曼这样重要的时空就已经允许进入内视界的观测者直接观测到奇点，从而违反强宇宙监督假设，怎么还会有人“不识时务”地提出强宇宙监督假设呢？这是因为有迹象表明像雷斯勒-诺斯特朗姆和克尔-纽曼时空中的内视界这样的东西很可能是不稳定的。这些内视界是所谓的柯西视界，它是强宇宙监督假设遭到破坏时必然会出现的^{[\(20\)](#)}。1982年，钱德拉塞卡（Subrahmanyan Chandrasekhar, 1910—1995）与哈特尔（James Hartle, 1939—）证明了雷斯勒-诺斯特朗姆时空的柯西视界是一个无限蓝移面，即来自远处的辐射在这一视界处会被无限蓝移，从而不仅会对试图穿越这一视界的任何观测者造成致命伤害，而且会严重干扰雷斯勒-诺斯特朗姆度规本身，导致柯西视界的不稳定。很多物理学家认为，类似的不稳定性也存在于更一般的克尔-纽曼时空中，其结果便是使得像诸如雷斯勒-诺斯特朗姆或克尔-纽曼度规所描述的那种奇点为类时的情形无法出现^{[\(21\)](#)}。当然，这些基于特例所做的研究距离证明强宇宙监督假设所需的普遍性还相差很远。

有关宇宙监督假设的研究还在进行。就笔者个人的看法而言，虽然我对霍金版的宇宙监督假设是否成立这样的纯技术问题并无先入之见，不过我并不看好导致宇宙监督假设的那种推理。物理学家们大都相信，奇点在广义相对论中的存在是广义相对论作为经典理论的局限之处。在一个完整的量子引力或更宏伟的所谓终极理论（theory of everything）中，那样的奇点是不会存在的。既然如此，我们有什么理由认为广义相对论需要以某种精巧的方式阻止事实上根本就不存在的奇点的观测效应

呢？就拿霍金那句“上帝憎恶裸奇点”来说吧，如果“上帝”创造的原本就是一个量子引力世界，在那里原本就不存在真正的奇点，那它又有什么理由要在广义相对论这样一个非量子、从而注定是片面的理论中“憎恶”一个在完整理论中根本就不存在的裸奇点呢？退一步讲，倘若“上帝”果真“憎恶”广义相对论中的裸奇点，为何又允许它出现在种种个例之中呢？导致宇宙监督假设的那种推理，无论是作为拟人化的“上帝”的喜好，还是作为单纯的物理分析，都显得不够自洽。从历史上看，在黑洞概念问世之初，英国物理学家爱丁顿曾认为大自然将会阻止黑洞出现，就像宇宙监督假设的支持者们认为大自然将会阻止裸奇点出现一样，结果他错了；在量子力学问世之初，爱因斯坦曾像霍金那样猜测过“上帝”的喜好，认为他老人家不会掷骰子，结果他也错了。历史将会在宇宙监督假设这里重演一遍，还是玩出新的花样？宇宙监督假设的最终命运将会如何？这是未来广义相对论研究中一个引人注目的悬念。

注释

[\[1\]](#) 如果宇宙在未来演化中出现整体性的大坍缩（从目前所知的宇宙学常数来看，这应该不会发生），那么大坍缩也将是一个裸奇点。不过这个裸奇点是未来奇点，也是与可观测宇宙有因果关联的时间本身的终结，算不上破坏广义相对论的预言能力。

[\[2\]](#) 不过，即便在非渐近平直时空中，只要存在观测意义上足够接近渐近平直条件的时空区域，黑洞的概念就仍可近似成立。但初学者常为之困惑的“宇宙本身是否是黑洞”的答案则是否定的。因为对于宇宙本身来说，哪怕在近似意义上也不存在定义黑洞所需的渐近平直区域，因此黑洞这一概念并不适用。

[\[3\]](#) 有些文献——比如沃尔德的著名教材《广义相对论》（General Relativity）——在某些表述中将渐近平直与强渐近可预测列为彼此独立的条件。但实际上强渐近可预测时空的定义（包括沃尔德采用的定义）往往首先就假定了渐近平直性。

[\[4\]](#) 强渐近可预测时空定义的严格表述会涉及一些微妙的数学细节，其中包括我们在3.1节中定义渐近平直时空时引进的“非物理时空”。此外，不同文献采用的定义有时带有细微差

别，这些我们就不展开讨论了。

[\[5\]](#)从理论上讲，奇点既可以存在于黑洞内部，也可能存在于视界上（可以证明黑洞是闭区域，视界是黑洞的一部分）。后一种可能性通常被忽略（即假定视界是非奇异的），不过对于宇宙监督假设来说，我们只需强调时空的强渐近可预测性，奇点是否出现在视界上对讨论没有影响。

[\[6\]](#)克尔-纽曼度规有时也被简称为克尔度规。不过克尔度规有时又专指1963年克尔发现的不带电的稳定轴对称真空解。对于这些命名上的细微差异，读者在阅读文献时需按照上下文进行判断。

[\[7\]](#)需要注意的是，在这些考虑中引力辐射及入射粒子对背景时空的干扰通常被忽略了。此外，文献中似乎未见有人分析过向极端带电黑洞投放带角动量的粒子，或向带有极端角动量的黑洞投放电荷的情形。沃尔德曾发表过一篇题为“摧毁黑洞的理想实验”（*Gedanken Experiments to Destroy a Black Hole*）的文章，对向极端克尔-纽曼黑洞投放电荷的情形进行系统分析。但他的分析看似普遍，却并不适用于黑洞带极端角动量（即 $J=m$ ， $Q=0$ ）的特殊情形。另外值得一提的是，对背景时空为渐近de Sitter时空的情形所做的类似研究表明，在那种情形下通过向黑洞投放电荷或角动量来摧毁黑洞是可以做到的。不过这对我们的讨论并不产生影响，因为黑洞的定义、时空的全局性质等，本身就与背景时空的渐近结构有着密切关系，不同的背景时空所对应的物理问题是完全不同的。

[\[8\]](#)这一结果与所谓的黑洞热力学第三定律也有很好的相容性，因为按照黑洞热力学，极端黑洞的温度为零，而黑洞热力学第三定律表明无法使黑洞的温度达到绝对零度，即无法使黑洞达到极端状态。

[\[9\]](#)彭罗斯最初考虑的是球对称的辐射物质，也叫做零尘埃（*null dust*）——即能量动量张量为 $\rho k^\mu k^\nu$ （ k^μ 为类光矢量）的物质。但球对称与辐射物质这两个条件对论述来说并不是必需的。

[\[10\]](#)有些文献将证明黑洞面积定理所需的能量条件表述为弱能量条件，甚至强能量条件，这是不必要的要求。

[\[11\]](#)只能说是“基本被排除”而不是彻底被排除，因为并不存在关于这一排除的严格证明，而且直到近期仍有个别物理学家在进行这方面的努力。

[\[12\]](#) 伯克霍夫定理的一种早期，并且直到现在依然很流行的表述形式是“真空中的球对称度规必定是静态度规”。不过后来人们发现施瓦西度规在延拓区域内不一定是静态的，因此“真空中的球对称度规必定是施瓦西度规”被认为是更准确的表述。

[\[13\]](#) 严格地讲，使该奇点成为裸奇点还要求当施瓦西视界形成时，从奇点发出的光信号已经传播到视界之外（请读者想一想，这是为什么？）。

[\[14\]](#) 那些条件全都是在满足适当的能量条件——比如主能量条件——的基础之上提出的。

[\[15\]](#) 壳层会聚奇点既可以位于有限半径处，也可以位于坍缩球体的中心，不过位于有限半径处的壳层会聚奇点通常晚于视界而出现，从而不是裸奇点。

[\[16\]](#) 请读者想一想，理想流体为什么不满足这一定义？

[\[17\]](#) 柯西面的定义可参看2.4节，“最大柯西展开”的详细定义可参阅沃尔德的教材。

[\[18\]](#) 爱因斯坦-克莱因-戈登体系中的标量场一般来说是有质量的，本节涉及的零质量情形只是特例。

[\[19\]](#) 霍金新赌约中涉及“一般初始条件”的原文是“源自一般初始条件（即源自初始条件组成的开集）的动力学演化不可能产生裸奇点”。这一表述似乎只有从“能产生裸奇点的初始条件集不包含开子集”这一角度来理解才既有意义，又不与已知的反例相矛盾。如果能产生裸奇点的初始条件集不包含开子集，那么任何能产生裸奇点的初始条件都可以被某些微扰所破坏（请读者结合开集的定义自行论述这一点），从而是不稳定的。但另一方面，经过这样理解的“一般初始条件”并不意味着能产生裸奇点的初始条件是“稀有”的（请读者想一想，这是为什么？）。从而严格讲并不足以剥夺裸奇点的现实意义。可惜霍金似乎从未在学术论文中阐述过这一条件，从而使我们无法找到更可靠的参照。在笔者看来，对“一般初始条件”的更恰当的定义是要求能导致裸奇点的初始条件的相对测度为零。在这方面，宇宙监督假设的另一位核心支持者彭罗斯的看法要狡猾得多，他认为现阶段最好别给“一般初始条件”下明确定义，以便保留随机应变的能力。

[\[20\]](#) 这是因为强宇宙监督假设要求时空具有全局双曲性，因而它的破坏意味着初始条件的柯西展开具有边界，这一边界——如我们在2.4节中所介绍的——正是柯西视界。

[\[21\]](#) 需要注意的是，柯西视界在赖斯纳-诺斯特朗姆和克尔-纽曼情形下的不稳定性并不

意味着它在任何情形下都不稳定。一个引人注目的反例是：莫里斯（Mike Morris）、索恩和尤瑟福（Ulvi Yurtsever）曾在1988年发现，虫洞时空中的柯西视界不具有由无限蓝移导致的不稳定性。不过，如我们在虫洞物理学中将会看到的，虫洞时空并不满足宇宙监督假设所要求的能量条件。

第5章 虫洞物理学

5.1 萨根的小说

从本节起，我们将进入本书的最后一个专题：虫洞物理学。虫洞（wormhole）这个朴素的术语会让人产生一些朴素的联想，比如联想到虫子在苹果上咬出的洞。它们在拓扑学上的含义也确实类似，只不过那“苹果”要换成时空，而那“虫子”有可能是我们自己。如果联想得更多一点的话，还可能由拓扑而联想到另一个层面上，那就是度量，或者说长度。苹果中的虫洞若足够直的话，是可以成为连接苹果表面两个点的捷径的——即比沿苹果表面的路径有更短的长度。物理学家希望时空中的虫洞也能如此，即成为连接两个时空区域的捷径。

那样的捷径对于时常要用自己的想象力去触碰光速极限的科幻作家来说，无疑是便利的工具和巨大的福音。不过，与科幻作家从科学概念中寻找工具的老套故事不同的是，在虫洞这一工具的问世过程中，科幻作家本身就起了一定的促进作用，这一点是足可令科幻作家们自豪的。唯一使这自豪稍稍逊色的，是那起了促进作用的科幻作家乃是业余的，或者说只是一位客串者——他的正式身份是天文学家。

1985年，美国康奈尔大学（Cornell University）的天文学家萨根（Carl Sagan, 1934—1996）发表了一部很著名、并在12年后（即1997年）被拍成电影的科幻小说，书名叫做《接触》（Contact）。在这部小说中，包括女主人公在内的五位地球人（在电影中改为了女主人公一人）通过以外星智慧生物传来的信息为蓝图建造起来的装置，在无论对于他们自己还是对于其他人来说都很短的时间内，旅行到了距地球26光年的织女星（Vega）附近，与外星文明进行了接触，并顺利返回了地

球。

小说中的人物是如何跨越广袤的恒星际空间的呢？聪明的萨根让他们自己进行了讨论，讨论的结果是认为他们的旅行是通过虫洞进行的。萨根并且采用了我们上面提到的有关苹果的比喻，表示——当然，仅限于他的小说之中——宇宙就像一个大苹果，我们人类生活在它的表面上，而外星智慧生物则建造了穿越苹果内部的纵横交错的虫洞，作为星际旅行的通道。

虽然只是科幻世界的客串者，但萨根所具有的天文学家与科普作家的双重身份，以及此前不久因电视科普系列片《宇宙》（Cosmos）而获得的公众知名度，使他的小说产生了巨大影响。事实上，他的小说尚未动笔就已获得了200万美元的据说是史上最高的预付稿费。而小说发表之后，他所采用的虫洞这一概念，则很快风靡科幻界，成为了其他科幻作家竞相追随的标准概念，以及星际旅行的标准手段。

在撰写小说的那段时间里，萨根住在纽约上州一栋颇具乡村风格的屋子里。周围是山丘地带，地势起伏不定，还间杂了峡谷、沟壑、悬崖、瀑布等。美国科普作家哈尔彭（Paul Halpern）在一本有关虫洞的科普作品中曾饶有兴致地罗列了那些地形，并风趣地表示那些都是诱使人们思考时空结构的理想地形。确实，在那样复杂多变的地形下，人们很少能沿直线从一处旅行到另一处，要想有快捷的旅行，就得在某些地方构建像虫洞那样的通道。不过，地形虽然理想，萨根恐怕却并未从中得到过有关时空结构的灵感，起码他小说中的虫洞概念并非受地形启发，而是另有缘由的。事实上，尽管萨根对于虫洞概念的流行甚至兴起都有着促进之功，最主要的功劳却必须归于他的一位老朋友——加州理工学院（California Institute of Technology, Caltech）的物理学家索恩。

那是虫洞物理学中的一段佳话，索恩本人曾就此写过一段回忆：

我刚刚教完1984—1985学年的最后一堂课，沉陷在办公室的椅子上让激情退去，这时电话

铃声响了起来。打来电话的是我的老朋友、康奈尔大学的天体物理学家萨根。“很抱歉打扰你，索恩。”他说：“我刚刚写完一部有关人类与外星文明首次接触的小说，但却有些顾虑。我希望其中的科学尽可能地准确，可我担心自己也许会搞错某些引力理论的东西。你愿意看一下并给我一些建议吗？”我当然愿意。那将是有趣的，甚至有可能是好玩的，因为萨根是个聪明的家伙。更何况，我怎能拒绝这种来自朋友的要求呢？

几星期后，萨根寄来了一叠厚厚的书稿。当时索恩正要与前妻及儿子驱车前往女儿就读的大学参加毕业典礼，便带上了书稿。一路上，索恩的前妻和儿子轮流开车，索恩自己则将书稿浏览了一遍。“小说很有趣，但萨根确实遇到了麻烦。”——这是索恩读完书稿后的感觉。为什么会有这种感觉呢？因为萨根在书稿中所设想的星际旅行是通过黑洞进行的，而在索恩这样的引力理论专家眼里那是一条死路。在参加完毕业典礼返家的途中，当车子行驶在平坦的州际公路上时，索恩开始为萨根出谋划策。经过一番思考，他想到了十几年前由他的导师、美国物理学家惠勒提出的虫洞概念⁽¹⁾。这一概念当时还知者甚少，但索恩无疑是例外。这一概念可以帮助萨根吗？通过虫洞有可能进行星际旅行吗？索恩当即拿出理论物理学家的看家“法宝”——纸和笔——进行了分析，并得到了一个喜忧参半的初步结果（我们将在后文中详细介绍）。尽管只是喜忧参半，比起黑洞那死路一条来终究是乐观多了。回到家后，索恩便将通过虫洞进行星际旅行的设想及初步的理论分析告诉了萨根，后者接受了他的设想，于是就有了出现在《接触》这部小说、后来的同名影片，以及许许多多其他科幻作品中的那些通过虫洞进行星际旅行的精彩故事。

在向萨根推荐了通过虫洞进行星际旅行的设想之后，索恩自己也对虫洞产生了浓厚兴趣，与他的学生莫里斯一起，开始对虫洞物理学展开认真研究，并于两年后（即1987年）发布了结果。索恩和莫里斯的研究是对虫洞结构的第一次细致研究（那之前的虫洞基本上只是一个概

念），从而堪称是虫洞物理学的开山之作。不过，索恩和莫里斯自己对他们的研究做了一个很谦虚的定位，只定位为讲授广义相对论的教学工具。这可以从他们论文的标题中看出来，那标题是“时空中的虫洞及它们在星际旅行中的应用：讲授广义相对论的工具”（Wormhole in spacetime and their use for interstellar travel: A tool for teaching general relativity）。与这一定位相一致的是他们为论文所选的刊物：《美国物理杂志》（American Journal of Physics），那是一份著名的物理教学刊物。

5.2 黑洞、白洞和虫洞

不过，在介绍索恩和莫里斯的研究之前，让我们先对使索恩觉得萨根“确实遇到了麻烦”的设想——即通过黑洞进行星际旅行的设想——作一个简单的介绍及评述。这不仅是因为对一个“此路不通”的设想的正式排除，更能显出虫洞的意义，以及它在现代科幻作品中快速风靡的原因，而且也是因为那“此路不通”的设想遭遇失败的地方，恰恰孕育着虫洞物理学崛起的关键，从而对那些失败的了解，有助于我们更好地理解虫洞物理学。

初看起来，通过像黑洞那样恶名远扬，被认为能吞噬一切（其中当然包括宇航员）的危险天体来进行星际旅行是一种离奇的设想。不过，结论不能下得太快，因为萨根毕竟不是普通的科幻作家，能被他采用的设想除了离奇之外，多少是有些理论渊源的。事实上，那设想不仅有理论渊源，而且那渊源还非同一般，因为它来自于广义相对论的创始人——爱因斯坦。

1935年，爱因斯坦与以色列物理学家罗森（Nathan Rosen, 1909—1995）合作发表了一篇题为“广义相对论中的粒子问题”（The Particle Problem in the General Theory of Relativity）的论文。在那篇论文中，爱因斯坦和罗森提出了后来被称为爱因斯坦-罗森桥（Einstein-Rosen bridge）的广义相对论的特殊解。

什么是爱因斯坦-罗森桥呢？它有若干个版本，其中最简单的版本是对施瓦西度规

$$ds^2 = \left(1 - \frac{2m}{r}\right) dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2 d\Omega^2 \quad (5.2.1)$$

作坐标变换 $r - 2m = u^2$ 所得到的解（请读者自行验证）：

$$ds^2 = \frac{u^2}{u^2 + 2m} dt^2 - 4(u^2 + 2m) du^2 - (u^2 + 2m)^2 d\Omega^2 \quad (5.2.2)$$

这个解在 $u \in (-\infty, \infty)$ 的整个取值范围内都不包含 $r < 2m$ 的区域（从而避免了 $r=0$ 的奇点），且在 $u \rightarrow -\infty$ 和 $u \rightarrow \infty$ 时都是渐近平直的（因为对应于 $r \rightarrow \infty$ ），从而像是一座连接 $u \rightarrow -\infty$ 和 $u \rightarrow \infty$ 这两个渐近平直时空的“桥梁”。既然有“桥梁”，自然可以联想到旅行。不过，爱因斯坦和罗森提出这个解的用意与星际旅行是毫不相干的，他们的用意——如他们文章的标题所明示的——乃是用场的概念来表示粒子。这种试图将场和粒子统一起来（确切地说是统一在场的框架内）的努力是爱因斯坦晚年研究的核心课题之一，也是他寻求统一理论的一个重要环节。上述爱因斯坦-罗森桥在爱因斯坦眼里乃是广义相对论的一个处处有限（因为避免了 $r=0$ 的奇点）、有可能用来表示中性粒子的特殊解。

爱因斯坦的这一努力与他晚年在寻求统一理论之路上的其他尝试一样，并未取得成功。但爱因斯坦-罗森桥所具有的连接两个渐近平直时空的特性，起码在表观上与虫洞有着异曲同工之处⁽²⁾。由于上述爱因斯坦-罗森桥是从描述施瓦西黑洞的施瓦西度规中构造出来的，因此可在一定程度上被视为是通过黑洞进行星际旅行这一设想的理论渊源。

不过，表观上的异曲同工并不等于实质上相似，渊源的非同一般也并不意味着设想可信。如果你试图——像萨根原先设想的那样——通过上述爱因斯坦-罗森桥进行星际旅行，你的结局将会是很悲惨的，因为你将不会经过任何“桥梁”，更不会到达什么渐近平直时空，而是会直接落入施瓦西黑洞。爱因斯坦和罗森通过本质上是将黑洞外部区域覆盖两次的特殊坐标变换构造出的所谓“桥梁”，其实只是坐标缺陷带来的幻象，因为在施瓦西视界（ $r=2m$ 或 $u=0$ ）处，爱因斯坦-罗森桥的度规是退化的（因为 $g_{00}=0$ ）。除上述最简单的版本外，爱因斯坦-罗森桥还有更复杂的版本，被爱因斯坦和罗森用来试图表示带电粒子。那种版本的问题更多，除同样存在坐标缺陷外，还有其他方面的问题，就不赘述

了。

而更糟糕的是，即便“视死如归”地撇开上述所有问题，通过爱因斯坦-罗森桥进行星际旅行也还存在其他麻烦，比如众所周知（感兴趣的读者可以自行验证），物体抵达施瓦西视界（即抵达上述爱因斯坦-罗森桥的正中央）的过程在外部观测者看来是要花费无穷长时间的。这使得通过上述爱因斯坦-罗森桥进行星际旅行即便可以“苟全性命”，也只对旅行者本人才有意义，外部观测者却是永远也无法看见旅行过程的完成，更不可能从中受益（哪怕是信息意义上的受益）的。

因此，爱因斯坦-罗森桥是不能够作为星际旅行的通道的。

不过，这并不等于通过黑洞进行星际旅行就被排除了，因为爱因斯坦-罗森桥只是通过黑洞进行星际旅行的若干设想之一。这方面的另一类设想，是通过广义相对论的某些更复杂的解（比如雷斯勒-诺斯特朗姆解或克尔-纽曼解）的某些奇妙特点（比如奇点成为类时甚至变成奇环，从而可被绕开或穿越）进行星际旅行。但是，这类设想也面临一些棘手的问题，比如我们在4.7节中曾经提到过，雷斯勒-诺斯特朗姆解和克尔-纽曼解的内视界很可能是所谓的无限蓝移面，它会对任何试图穿越或靠近的旅行者造成致命伤害，甚至破坏雷斯勒-诺斯特朗姆解和克尔-纽曼解本身。此外，黑洞引力场的不均匀性所带来的所谓潮汐力（tidal force）也很可能会对旅行者产生致命伤害^[3]。

除上述种种麻烦外，通过黑洞进行星际旅行的设想还面临另外一个棘手的问题，那就是出口问题。黑洞作为吞噬一切的家伙，进入它不是问题，但进入之后不管从什么地方出来，都似乎有悖其“吞噬一切”的恶名。这个问题该如何解决呢？这就引出了一个或许算得上是引力理论中最离奇的概念——白洞（white hole）。

顾名思义，白洞就是性质与黑洞完全相反的天体：黑洞只进不出，白洞就只出不进；黑洞吞噬一切，白洞就喷射一切。那样的天体如果存在，就有可能作为通过黑洞进行星际旅行时的出口。但与黑洞之顺应引

力的基本性质，从而受到理论家们的普遍青睐，甚至获得了较强的间接观测支持不同，白洞的喷射乃是逆引力而行，从而依赖于极为离奇的初始条件。事实上，白洞这一概念之所以存在，唯一的理由就是作为黑洞的时间反演——因为广义相对论具有时间反演对称性。但我们都知道，很多原则上可以存在的时间反演过程在实际上是几乎绝不可能出现的。比如你扔一块石头到水里，发出“扑通”的声音，并荡起一圈圈向外延展的涟漪，这个过程的时间反演，即涟漪和声波向内收缩，将石头从水里反弹到你手里的过程，虽然原则上是可以存在的，但实际上却几乎绝不可能出现。白洞作为黑洞的时间反演也是如此。广义相对论的黑洞解所具有的极大的简单性也许会给人一个错觉，以为黑洞的时间反演未必罕见，但实际上，若果真要反演一个黑洞，还必须反演它的蒸发过程，仅此一点，就足以将概率降低到难以想象的小。更何况，倘若要反演物体落入黑洞的过程，就还必须反演这一过程中发射的各种辐射（包括引力辐射）等，概率之小就更是无法形容了。不仅如此，所有跟黑洞有关的麻烦，在白洞上也会有一定程度的体现。比如拿最简单的施瓦西黑洞来说，物体抵达视界的过程在外部观测者看来需要无穷长的时间，相应的，假如有白洞的话，物体从视界中喷射出来的过程在外部观测者看来也将需要无穷长的时间，或者换句话说，喷射过程在外部观测者看来应该是从无穷长时间之前就开始了，这在可观测宇宙中是根本不可能实现的。因此，白洞的背后即便不是彻底的死路，起码也是希望极为渺茫的，顺带着也给通过黑洞进行星际旅行的设想又泼了一大瓢冷水。

黑洞、白洞都靠边站了，自然就轮到虫洞出场了[\(4\)](#)。虫洞这一概念——如前所述——乃是出自索恩的导师、美国物理学家惠勒之手。20世纪50年代，惠勒展开了对所谓“几何动力学”（*geometrodynamics*）这一新兴、可惜始终未能真正兴起的领域的研究。这一研究的核心目标是将物理学几何化。为了达到这一目标，一个必不可少的步骤就是将粒子（包括带电粒子）几何化。细心的读者也许看出来了，惠勒的这一目标

跟爱因斯坦和罗森研究爱因斯坦-罗森桥的用意十分相似。事实上，上面所介绍的最简单的爱因斯坦-罗森桥正是试图用已经几何化了的引力场来表示粒子，从而可以视为将粒子几何化的早期尝试。而惠勒的工作则是继爱因斯坦-罗森桥之后，这一领域内20年间几乎唯一重要的新尝试。

在这种新尝试中，惠勒引进了物理空间的多连通结构，这种多连通结构的一个简单图示出现在他发表于1955年的题为“几何子”（Geons）的论文中，正是典型的虫洞图示（见图5.1）。在这一图示中，惠勒画出了一些电磁场的力线，它们从虫洞的一个“嘴巴”（mouth）进入，又从另一个“嘴巴”出去，无须任何电荷就显示出了如同一对正负电荷所产生的电磁场。两年后（即1957年），惠勒与自己的学生米斯纳在一篇题为“作为几何学的经典物理学”（Classical physics as geometry）的论文中，将上述无须任何电荷就能显示出如同一对正负电荷所产生的电磁场的巧妙结果称为“没有电荷的电荷”（charge without charge）。类似地，他还提出了“没有质量的质量”（mass without mass），以及“没有场方程的场方程”（field equations without field equations）等颇带哲理气息的概念。这些都是他那几何动力学中的主要结果。除这些结果外，惠勒与米斯纳的那篇1957年的论文，也是虫洞这一术语的诞生地。惠勒与米斯纳在文章中表示，拓扑学家们会把他们提出的结构称为“多连通空间”（multiply-connected space），但物理学家们也许可以将之更生动地命名为虫洞——这正是索恩所继承，并经萨根普及出去的命名。

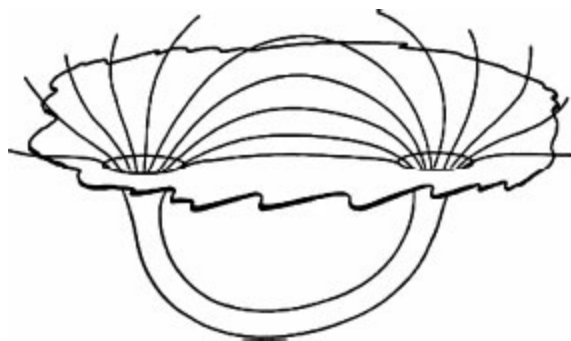


图5.1 惠勒的虫洞

与上述想法相匹配，惠勒还提出了有关时空结构的量子涨落（quantum fluctuation）的想法。按照这种想法，时空本身也像其他物理体系一样，在小尺度上会出现量子涨落。在极端情形下，比如当尺度小于所谓的“普朗克长度”（Planck length）时⁽⁵⁾，量子涨落将会变得如此巨大，就连时空的拓扑结构都有可能被改变，甚至有可能裂成碎片，出现所谓的“时空泡沫”（spacetime foam）。如果作个比喻的话，那么时空就像海面，从大尺度上看平滑如镜，随着尺度的缩小渐渐显出起伏，当尺度小到一定程度时，则可以看到汹涌的波涛乃至飞散的泡沫。惠勒的这种想法与包括虫洞在内的他那几何动力学中的若干基本概念是密切相关的。事实上，时空结构的这种量子涨落——如果存在的话——无疑可以成为虫洞出现的温床。

不过，惠勒所设想的时空结构的量子涨落从单纯类比的角度看虽颇为自然，并且受到了物理学家们迄今尚未建立量子引力理论这一不幸事实的自动“保护”，但许多物理学家对之仍是深深怀疑的，对此将在后文中作进一步介绍。另一方面，惠勒的上述想法所涉及的包括虫洞在内的各种时空结构基本上都是微观的，这对于他所追求的将粒子几何化的目标来说或许是足够了，但对于我们——以及萨根和索恩等人——感兴趣的星际旅行的需要来说却无疑差得很远。而且惠勒虽然提出了虫洞概念，但基本上只停留在概念层面，而并未对诸如虫洞的稳定性之类的技术细节给予足够关注，可那些技术细节对于通过虫洞进行星际旅行来说，却有着举足轻重的重要性。那些技术细节正是索恩和莫里斯的研究重点。因此，惠勒虽然是虫洞研究的先驱者，对虫洞物理学进行细致研究的荣誉却要归于索恩和莫里斯。索恩和莫里斯所研究的可作为星际旅行通道的虫洞有一个专门的名称，叫做“可穿越虫洞”（traversable wormhole）。索恩和莫里斯是可穿越虫洞研究的开创者。

5.3 球对称可穿越虫洞

可穿越虫洞与惠勒提出的概念层面上的虫洞的差别就在于“可穿越”三个字。究竟什么样的虫洞是可穿越的呢？这是索恩和莫里斯首先要确定的。我们在上节中介绍通过黑洞进行星际旅行的设想时曾经表示，该设想遭遇失败的地方，恰恰孕育着虫洞物理学崛起的关键。现在就让我们来盘点一下通过黑洞进行星际旅行的失败之处。

通过黑洞进行星际旅行的最核心的失败之处显然就在于存在视界，它所导致的困难是多重的：比如它的“只进不出”要靠白洞那样的离奇概念来“解套”；比如落入或离开它的过程在外部观测者看来要花费无穷长的时间；比如它有可能是致命的无限蓝移面。因此，可穿越虫洞所需满足的首要条件就是不存在视界。

通过黑洞进行星际旅行的另一个失败之处是引力场的不均匀性造成的潮汐力，虽然如我们在上节的注释中所说，这个失败之处并没有通常渲染的那样严峻，但它无疑是可穿越虫洞必须“引以为戒”的。因此，可穿越虫洞所需满足的另一个条件是穿越过程中遇到的潮汐力是人体能够承受的。考虑到潮汐力未必是穿越过程中有可能遇到的唯一应力，更普遍的条件可以表述为穿越过程中遇到的应力是人体能够承受的。

这两条就是从通过黑洞进行星际旅行的失败中得到的“经验教训”。除此之外，可穿越虫洞还必须满足一些一般性的理论条件：首先是它必须满足广义相对论场方程；其次是它的物质分布必须是物理上可以实现的——这包括物质的能量动量张量是物理上存在的，以及物质的数量是可观测宇宙可以提供的；最后则是它必须能在微扰下保持稳定——否则的话，星际飞船通过时带来的干扰就有可能破坏可穿越虫洞。

这些就是索恩和莫里斯归纳出的可穿越虫洞所需满足的条件。不过这些条件都很一般，为了便于具体计算，他们还引进了一些简化条件：

首先是假设了可穿越虫洞的度规是静态球对称的。这当然不是必需的，但在广义相对论研究中乃是首选的简化条件，比如广义相对论的第一个严格解——施瓦西解——就是在这一简化条件下得到的，从它入手进行可穿越虫洞研究也是顺理成章的。而且从物理上讲，虫洞如果是一种大尺度物质结构，它的天然形态也确实有可能像其他大型天体一样是接近静态球对称的。其次是假设了可穿越虫洞的所谓“喉咙”（throat）——即径向坐标值的最小处——是唯一的，或者换句话说，径向坐标 r 作为径向本征距离 $s = \int dr$ 的函数有唯一的最小值 r_0 。这当然也不是必需的，因为虫洞的“喉咙”完全可以是更复杂的。不过如我们将会看到的，“喉咙”是虫洞性质最独特的地方，因此对它的简化是很有帮助的。最后则是假设了可穿越虫洞的出入口分别连接渐近平直时空。这同样也不是必需的，因为在非渐近平直时空中也可以有虫洞。但正如在非渐近平直时空中可以存在黑洞那样的东西，物理学家们研究黑洞时仍普遍假设时空是渐近平直的，虫洞研究也是如此。这一简化条件还可以这样来理解，那就是虫洞本身的结构与时空的大尺度结构并无密切关系，因此不妨对后者采用最便利的假设^{[\(6\)](#)}。

为清楚起见，我们把上面提到的可穿越虫洞所需满足的所有条件罗列在一起^{[\(7\)](#)}：

- (1) [“经验教训”] 不存在视界；
- (2) [“经验教训”] 穿越过程中遇到的应力是人体能够承受的；
- (3) [一般条件] 满足广义相对论场方程；
- (4) [一般条件] 物质的能量动量张量是物理上存在的；
- (5) [一般条件] 物质的数量是可观测宇宙可以提供的；
- (6) [一般条件] 在微扰下保持稳定；
- (7) [简化条件] 度规是静态球对称的；
- (8) [简化条件] “喉咙”是唯一的；
- (9) [简化条件] 出入口分别连接渐近平直时空。

条件列出了，接下来就是寻找满足条件的具体虫洞解了。由于下面讨论的全都是可穿越虫洞，为行文简洁起见，有时将会略去“可穿越”这一限定词。在广义相对论中，寻找具体解的传统做法是首先给定物质分布（即物质能量动量张量的分布），然后求解广义相对论场方程以得到时空结构。这一做法体现的是物质为因、几何为果的物理思想，或者用惠勒的话说：“物质告诉时空如何弯曲，时空告诉物质如何运动。”不过对于虫洞来说，这种做法很不方便，因为虫洞的物质分布在索恩和莫里斯的研究之前乃是无人知晓的东西，倒是它的时空结构早在惠勒的概念性研究中就已经有了直观图示。因此，索恩和莫里斯采用了一个聪明的思路，那就是将传统做法逆转，即从时空结构入手，然后用广义相对论场方程计算出物质分布。这种逆转在数学上是完全等价的（颠过来倒过去都是广义相对论场方程），在物理上却有着微妙的差别，那就是传统做法由于首先给定了物质分布，因此可以直接保证物质分布是物理上可以实现的（即满足条件（4）和（5）），而逆转的做法却无法直接保证这一点。这一微妙差别导致的后果我们很快就会看到。

由于虫洞的出入口分别连接渐近平直时空（条件（9）），这启示我们引进两个坐标域（coordinate patch），分别描述出口和入口附近的时空，两者在“喉咙”处相互衔接。而度规因为是静态球对称的（条件（7）），其一般形式是众所周知的，包含两个任意函数，且两者都只是径向坐标 r 的函数。由于广义协变性，度规形式的选择有很大的自由度。对于我们来说，比较方便的做法是将之表述成与施瓦西度规有一定类似性的形式：

$$ds^2 = e^{2\varphi_{\pm}(r)} dt^2 - \left[1 - \frac{b_{\pm}(r)}{r} \right]^{-1} dr^2 - r^2 d\Omega^2 \quad (5.3.1)$$

其中 φ_{\pm} 和 b_{\pm} 的下标 \pm 分别表示两个坐标域，坐标 r 的取值范围是 $[r_0, \infty)$ ， r_0 是“喉咙”所对应的径向坐标值。由于广义相对论场方程（条件

(3)) 是二阶微分方程，因此我们要求 $\varphi_{\pm}(r)$ 和 $b_{\pm}(r)$ 起码是二次可微的。我们并且还要求 $\varphi_{\pm}(r)$ 处处有限，这是不存在视界（条件

(1)) 的体现，因为它保证了 $e^{2\varphi_{\pm}(r)}$ 不会像施瓦西度规那样在某些地方（即视界上）为零。

除上述一般限定外，这一度规在 $r=r_0$ 和 $r \rightarrow \infty$ 需要满足一些边界条件。对于 $r \rightarrow \infty$ ，由于出入口连接渐近平直时空（条件(9)），因此 $\varphi_+(\infty)$ 和 $\varphi_-(\infty)$ 均为（有限）常数。在一般情况下，这两个常数可以是不相等的（请读者想一想，这两个常数不相等的物理意义是什么？）。同样的， $b_+(\infty)$ 和 $b_-(\infty)$ 也均为（有限）常数。与施瓦西度规相对比不难看出， $b_+(\infty)$ 和 $b_-(\infty)$ 分别对应于在出入口所连接的渐近平直时空中测得的出入口——也称为“嘴巴”（mouth）——的质量（确切地说是质量的两倍）。一般来说，这两个常数也可以是不相等的，即虫洞的两个“嘴巴”的质量可以是不相等的⁽⁸⁾。

对于 $r=r_0$ （即“喉咙”处），由于两个坐标域在此衔接，且 $\varphi_{\pm}(r)$ 和 $b_{\pm}(r)$ 起码是二次可微的。因此 $\varphi_+(r_0) = \varphi_-(r_0)$ ， $\varphi'_+(r_0) = \varphi'_-(r_0)$ ， $b_+(r_0) = b_-(r_0)$ ， $b'_+(r_0) = b'_-(r_0)$ 。不仅如此， $r=r_0$ 作为虫洞的“喉咙”，是径向坐标取值最小的地方，因此此处沿径向的 $dr/ds = [1 - b_{\pm}(r)/r]^{1/2} = 0$ ，而 $d^2r/ds^2 = [b_{\pm}(r)/r - b'_{\pm}(r)] / (2r) \geq 0$ 。这表明（请读者自行证明）

$$\begin{cases} b_+(r_0) = b_-(r_0) = r_0 \\ b'_+(r_0) = b'_-(r_0) \leq 1 \end{cases} \quad (5.3.2)$$

由于“喉咙”是唯一的（条件(8)），因此在偏离但靠近“喉咙”的一个开区间 $(r_0, r_0 + \Delta)$ 内， $dr/ds > 0$ ， $d^2r/ds^2 > 0$ ，或者等价地（也请读者自行证明）：

$$\begin{cases} b_{\pm}(r) < r \\ b'_{\pm}(r) < \frac{b_{\pm}(r)}{r} \end{cases} \quad (5.3.3)$$

以上就是球对称可穿越虫洞的时空结构所需满足的一般条件。另一方面，度规的静态球对称（条件（7））也给物质能量动量张量的形式施加了一定的限制，使它在 t 、 r 及两个横向坐标组成的正交标架场中具有 $T^{ab} = \text{diag}(\rho, \tau, p, p)$ 的正则形式（请读者想一想，球对称体现在哪里？），其中 ρ 是能量密度， τ 是径向张力， p 是横向压强，它们都只是径向坐标 r 的函数。这是球对称可穿越虫洞的物质分布所需满足的一般条件。

下一步要做的就是所谓的“用广义相对论场方程计算出物质分布”，由于我们已将物质分布归结为 ρ 、 τ 、 p 这三个函数，因此这一步实质上就是用前面引进的球对称度规计算出相应的爱因斯坦张量，将之——依据广义相对论场方程（条件（3））——与 $8\pi T^{ab}$ 等同起来，从而得到描述时空结构的函数 φ 、 b 与描述物质分布的函数 ρ 、 τ 、 p 之间的关系（这里我们丢弃了 φ_{\pm} 和 b_{\pm} 中表示坐标域的下标 \pm ，因为接下来的计算与坐标域无关）。这一计算是直截了当的，因为静态球对称度规的爱因斯坦张量的计算是广义相对论中的标准内容（当然，具体计算与所选择的度规形式有关），结果也并不复杂（感兴趣的读者可以自己推算一下）：

$$\begin{cases} 8\pi\rho = \frac{b'}{r^2} \\ 8\pi\tau = \frac{2(r-b)\varphi'}{r^2} - \frac{b}{r^3} \end{cases} \quad (5.3.4)$$

其中比这两个方程更复杂的关于 p 的方程因后面不会用到而省略了。

由这两个方程可以得到两个重要结果。一个是在“喉咙”处，由上述第二个方程可知

$$\tau(r_0) = -\frac{1}{8\pi r_0^2} \quad (5.3.5)$$

这个结果对于探讨虫洞是否真的“可穿越”有着重要影响，我们将在后文中进一步讨论。另一个是将这两个方程相加：

$$8\pi(\rho + \tau) = \frac{b'}{r^2} + \frac{2(r-b)\varphi'}{r^2} - \frac{b}{r^3} = -\frac{e^{2\varphi}}{r} \left[e^{-2\varphi} \left(1 - \frac{b}{r} \right) \right]' \quad (5.3.6)$$

由此不难看出在“喉咙”处（请读者自行证明）：

$$\rho(r_0) + \tau(r_0) = -\frac{[b'(r_0) - b(r_0)/r_0]}{8\pi r_0^2} \leq 0 \quad (5.3.7)$$

将这一结果与1.2节所介绍的能量条件相比较，可以看到零能量条件已是岌岌可危了——若不是“ \leq ”中还包含有“ $=$ ”的话，就已经被破坏了。但这个“ $=$ ”管得了“喉咙”却管不了周围，因为式（5.3.6）中的 $e^{-2\varphi}(1-b/r)$ 在“喉咙”处为零⁽⁹⁾，在偏离但靠近“喉咙”的一个开区间 $(r_0, r_0 + \Delta)$ 内却大于零（参阅式（5.3.3）），因此在偏离但靠近“喉咙”的一个开区间 $(r_0, r_0 + \delta)$ 内， $[e^{-2\varphi}(1-b/r)]' > 0$ ，从而

$$\rho(r) + \tau(r) < 0 \quad (5.3.8)$$

这表明在球对称可穿越虫洞的“喉咙”附近零能量条件会遭到破坏。由于零能量条件比弱能量条件、强能量条件、主能量条件都弱，因此它的破坏意味着在球对称可穿越虫洞的“喉咙”附近弱能量条件、强能量条件和主能量条件都会遭到破坏。这是索恩和莫里斯所得到的最重要的结果之一，也是我们在5.1节中提到的索恩在参加完毕业典礼返家途中得到的“喜忧参半的初步结果”中“忧”的部分。

我们前面提到过，索恩和莫里斯在寻找虫洞解的过程中对传统做法的逆转从数学上讲是等价的，在物理上却有着微妙差别，现在看到的正是这一微妙差别导致的后果。不过，这一后果虽有可“忧”之处，却也恰恰是索恩和莫里斯逆转传统做法的价值所在，否则的话——即首先

给定物质分布的话，是很难想到要引入违反能量条件的物质分布的，从而也就得不到球对称可穿越虫洞解了。至于违反能量条件是否意味着物质分布不再是物理上可以实现的（即违反条件（4）和（5）），我们在后文中将会进一步讨论。

关于上面这个结果，还有一个有趣的尾声可以提一下。索恩和莫里斯的论文提交后不久，索恩以前的学生、当时在美国宾夕法尼亚州立大学（Pennsylvania State University）物理系任教的佩奇（Don Page）给莫里斯写了一封信，提到索恩和莫里斯的上述结果可以很容易地从霍金和艾里斯（George Ellis）的名著《时空的大尺度结构》（The Large Scale Structure of Space-Time）的某些结果中推得。索恩和霍金是老朋友兼“赌友”（我们在4.1节末尾曾经介绍过他们打过的一个赌），彼此是很熟悉的，但索恩对霍金所擅长、并在那部名著中详加运用的所谓“全局方法”却并不熟悉。昔日学生佩奇——他同时也是霍金的学生——的这封来信使索恩既震动又惭愧，他在“检讨”这件事情时“沉痛”地表示：“我感到多么愚蠢啊，我从未深入学习过全局方法（霍金和艾里斯书中的课题），现在我为此付出了代价。”

佩奇提到的方法确实要比索恩和莫里斯的容易许多。当然，这种“容易”是见仁见智的，因为那是“站在巨人的肩膀”（全局方法）上的容易。如果不熟悉全局方法，那它非但不容易，反而是相当困难的。不过，虽不曾强调过，我们在前面介绍奇点定理时所涉及的很多内容其实就属于全局方法。因此，我们恰好幸运地具备了“站在巨人的肩膀”上的便利。利用这种便利，让我们对佩奇提到的方法作一个简单介绍。这种方法的切入点是2.3节中的式（2.3.1）（即雷查德利方程）。这是描述测地线束的方程。对于我们的目的来说，测地线束中的测地线要选为类光测地线，固有时间 τ 要改为仿射参数 λ ，切矢量的记号 V 则要改为表示类光矢量的 k ，即

$$\frac{d\theta}{d\lambda} = -R_{ab}k^ak^b - \frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab} \quad (5.3.9)$$

可以证明，对于沿径向传播的类光测地线束，切变张量 σ_{ab} 恒为零，因此上式可以简化为 $d\theta / d\lambda = -R_{ab}k^ak^b - (1/3)\theta^2$ 。由于膨胀标量 θ 描述的是测地线束的会聚或发散，因而在“喉咙”处，即测地线束从会聚转为发散的地方， $\theta=0$ 。另一方面，由于自喉咙向外测地线束是发散的，即 $\theta > 0$ ，因此在“喉咙”附近的一个小区域内 $d\theta / d\lambda > 0$ ，从而 $R_{ab}k^ak^b < 0$ 。而这等价于 $T^{ab}k_ak_b < 0$ （请读者自行证明），即破坏了零能量条件——与索恩和莫里斯的结果相同。

佩奇提到的全局方法不仅“容易”，而且还有一个更重要的优点，那就是不涉及度规的具体形式，从而可以被推广到更普遍——比如没有对称性——的情形。事实上，利用这类方法，物理学家们已经证明了远比索恩和莫里斯的结果普遍得多的结果，比如：全局双曲时空中的可穿越虫洞至少会在一条类光测地线上破坏零能量条件（从而也破坏弱能量条件、强能量条件和主能量条件）。由于全局双曲是一种非常优良、因而在一定程度上被认为是物理时空所具有的品质（虽然这种品质从理论上讲似乎太强了一点，而且极难有验证的可能）。因此这一结果表明可穿越虫洞对零能量条件（以及弱能量条件、强能量条件和主能量条件）的破坏很可能是普遍的[\(10\)](#)。

5.4 奇异物质——负能量的挑战

在上节中，我们得到了一个很可能适用于所有可穿越虫洞的普遍结论，那就是可穿越虫洞的物质分布（即物质能量动量张量的分布）会破坏零能量条件（从而也破坏弱能量条件、强能量条件和主能量条件）。物理学家们把那样的物质称为“奇异物质”（exotic matter）[\(11\)](#)。很明显，奇异物质之所以“奇异”，就在于零能量条件遭到了破坏。但是，零能量条件的破坏本身到底又“奇异”在哪里呢？对普通读者来说恐怕不是一清二楚的，因为能量条件哪怕对于物理专业的人来说，也并不属于“日常用语”。在本节中，我们首先要做的，就是将零能量条件的破坏具体化，将它与物理专业乃至普通人的“日常用语”联系起来，以显示其“奇异”之所在。

我们知道（参阅1.2节），零能量条件要求对于任何一个主压强 p ， $\rho + p \geq 0$ 。因此，零能量条件的破坏意味着起码对于某个主压强 p ， $\rho + p < 0$ 。为了显示这一条件的物理实质，让我们引进一个沿 p 所代表的主方向以速度 v 运动的惯性系 Σ 。由能量动量张量 $T^{ab} = \text{diag}(\rho, p)$ （这里我们略去了对当前计算没有影响的两个空间维度）的洛伦兹变换可以得到 Σ 系中的能量密度为（感兴趣的读者请自行计算一下）

$$\rho' = T^{00'} = T^{ab} \Lambda_a^0 \Lambda_b^{0'} = \gamma^2 (\rho + p) - p \quad (5.4.1)$$

其中 Λ 为洛伦兹变换矩阵， $\gamma = (1 - v^2)^{-1/2}$ 是洛伦兹因子（Lorentz factor）。由于 $\rho + p < 0$ ，而 γ 可以任意大（只要速度足够接近光速），因此总可以选择惯性系 Σ ，使得 $\rho' < 0$ ，即总可以选择惯性系 Σ ，使得其中观测到的能量密度小于零！这种负能量密度就是零能量条件的破坏所具有的“奇异”性质。由于这一性质，奇异物质也被称为负能量物质。

如果说得更浅白一点的话，那么奇异物质之所以“奇异”，是因为在

物理学中，能量的零点是用真空来定义的。相对于这一零点，任何其他状态（即有物质的状态）都具有正能量，物质越多，能量就越高；物质越少，能量则越低。依照这样的定义，负能量作为比零能量更低的能量，意味着比号称一无所有的真空具有“更少”的物质，这在经典物理学中是不可思议的，甚至近乎于语义上矛盾的说法。

假如这种“不可思议”或“语义上矛盾的说法”就是故事的终点，那么通过虫洞进行星际旅行或所谓“可穿越虫洞”的美丽设想也就可以算走到终点了。幸运的是，这种“不可思议”或“矛盾”来自于真空“一无所有”这样一个经典物理学的观念，而我们知道，经典物理学并不是物理学的终点，在它之后还有所谓的量子理论。量子理论的出现修正甚至颠覆了许多经典物理学的观念。那么，使负能量陷入困境的真空“一无所有”的观念是否也在此列呢？答案是肯定的，因为量子理论的发展彻底改变了我们对真空的理解。在量子理论中，真空所表示的乃是量子场的基态，它与其他状态（即有物质的状态）的区别，远没有像经典物理学中的“无”和“有”的区别那样截然。从某种意义上讲，量子理论中真空与其他状态的区别，不过是量子场能级的“低”和“高”之间的区别而已，定量多过定性。真空在量子理论中不仅像其他状态一样有自己的结构，而且还是高度动态的，随时可以有虚粒子对的产生和湮灭。在这种全新的、不再“一无所有”的真空下，负能量至少从概念层面上讲不再是不可思议，更不是语义上矛盾的了。

当然，这只是为负能量扫清了概念障碍，而并不足以确立它的实际存在。负能量的实际存在——如果可以确立的话——必须诉诸实验，最低限度——假如实验一时无法实现的话——也需要有更具体、更直接的理论支持。这是负能量对虫洞物理学的基本挑战。幸运的是，这种挑战得到了很正面的回应。经过理论物理学家与实验物理学家长期而共同的努力，理论支持与实验确立先后得以实现。

我们先来看看理论支持。既然量子理论的真空是有结构的，那么不

难想象，这种结构与几乎所有其他物理体系的结构一样，会受边界条件的影响。这就启示我们考虑特定边界条件下的真空，假如那种真空的能量比“普通”（即边界可以忽略的）真空具有更低的能量，那么就可以被视为一种负能量。沿这一思路的最著名的结果，是荷兰物理学家卡西米尔（Hendrik Casimir, 1909—2000）提出，并以他名字命名的卡西米尔效应（Casimir effect）。1948年，卡西米尔发表了一篇题为“论两块理想导体板之间的吸引”（On the attraction between two perfectly conducting plates）的论文，对两块相互平行的理想导体板——以下将简称为平行导体板——之间的真空能量面密度（即单位面积导体板之间空间内的总真空能量）进行了计算。结果表明，那能量面密度确实是负的！更重要的是，那能量面密度与平行导体板的间距有关（这是可以预料的，因为当间距趋于无穷时，因边界条件而产生的能量面密度必须消失），从而意味着平行导体板之间会因这种能量的存在而产生一种相互作用力。这种相互作用力被称为卡西米尔力（Casimir force），它的存在为从实验上检验卡西米尔效应提供了直接途径。

平行导体板情形下卡西米尔效应的计算并不复杂，只不过是对有边界条件和没有边界条件下的量子场基态能量——即所谓的零点能（zero-point energy）——进行比较而已，并且已被一些标准量子场论教材所收录，就不在这里重复了[\[12\]](#)。不过，直接把结果罗列出来又显得太偷懒，因此我们决定采取“中庸之道”，介绍一种很有用的半定量方法：量纲分析法（dimensional analysis），用它来给出计算结果的主要部分。众所周知，理论物理所用的方法几乎清一色是数学方法，如果一定要从中找出一些具有“物理特色”的方法的话，那么量纲分析法或许就是其中之一。这一方法虽然也披着简单的数学外衣，但其核心是属于物理的，因为量纲是物理量特有的东西。量纲分析法主要由两个步骤组成：

（1）通过对物理意义的分析，列出结果有可能依赖的所有物理量和物理常数；

(2) 通过对量纲的比较，确定结果对那些物理量和物理常数的具体依赖关系。

通过这两个步骤，运气好的话，可以将结果确定到只差一个常数（比例系数）的程度。平行导体板情形下的卡西米尔效应就是如此。

在卡西米尔效应的理论研究中，我们通常假定平行导体板足够大，从而导体板——作为二维客体——的边缘可以忽略⁽¹³⁾。在这种近似下，结果（即真空能量面密度）所能依赖的物理量只有一个，那就是平行导体板的间距，我们用 d 来表示，其余的就都是物理常数了。由于导体板所施加的边界条件是针对电磁场的，因此卡西米尔效应所涉及的量子场是电磁场，零点能则是电磁场的零点能，这种零点能只依赖于两个物理常数：光速 c 和普朗克常数 \hbar ⁽¹⁴⁾。因此，平行导体板情形下卡西米尔效应中的真空能量面密度 U 只依赖于平行导体板的间距 d 、光速 c 和普朗克常数 \hbar ，即 $U=f(d, c, \hbar)$ 。这就完成了两个步骤中的第一个。

那么具体的依赖关系是什么呢？这就需要对量纲进行比较了。为了简洁起见，我们用 $[X]$ 表示物理量或物理常数 X 的量纲，用 M 、 L 和 T 分别表示质量、长度和时间的量纲，则

$$\begin{cases} [U] = MT^{-2} \\ [d] = L \\ [c] = LT^{-1} \\ [\hbar] = ML^2T^{-1} \end{cases} \quad (5.4.2)$$

很明显，唯一能让函数关系 $U=f(d, c, \hbar)$ 的左右两边具有相同量纲的函数形式是（请读者自行证明）

$$U \propto \frac{\hbar c}{d^3} \quad (5.4.3)$$

这就完成了第二个步骤。可惜，用量纲分析法无法得到比例系数（这是该方法被称为“半定量方法”的原因所在）。对零点能的具体计算表明，

比例系数为 $-\pi^2 / 720$ 。这样，我们就得到了真空能量面密度的确切表示式，即

$$U = -\frac{\pi^2 \hbar c}{720d^3} \quad (5.4.4)$$

相应的体密度和能量动量张量则为

$$\rho = \frac{U}{d} = -\frac{\pi^2 \hbar c}{720d^4} \quad (5.4.5)$$

和（假定平行导体板的法向为z方向）[\(15\)](#)

$$T^{ab} = -\frac{\pi^2 \hbar c}{720d^4} \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & 3 \end{pmatrix} \quad (5.4.6)$$

由能量面密度，立刻可以得到平行导体板之间沿法向的相互作用力的面密度（即作用于单位面积导体板上的力）为

$$F = -\frac{\partial U}{\partial d} = -\frac{\pi^2 \hbar c}{240d^4} \quad (5.4.7)$$

其中结果中的负号表示这种力是相互吸引的。

以上就是有关平行导体板情形下卡西米尔效应的主要结果。在卡西米尔效应的研究中，有一段小小的历史值得一提。比针对平行导体板情形的研究稍早，卡西米尔与荷兰物理学家玻德（Dirk Polder, 1919—2001）曾合作发表过一篇论文，以类似于推导分子间范德瓦尔斯力

（van der Waals force）的方法，在大距离极限下，对理想导体板与单个的原子或分子，以及原子或分子彼此之间的相互作用进行了研究。那项研究通常被视为研究卡西米尔效应的先导，它所采用的方法也被视为对卡西米尔效应的另一种理解[\(16\)](#)。在那项先导研究的进行期间，卡西

米尔有一次访问哥本哈根（Copenhagen）时遇到了著名丹麦物理学家玻尔（Niels Bohr, 1885—1962）。玻尔问了一句物理学家们见面时的“标准问候”：你最近在研究什么？卡西米尔就将他与玻德正在进行的研究告诉了玻尔，并表示自己希望能用更简单、更优美的方法来进行推导。玻尔思考了一下，然后——依照卡西米尔的回忆——嘟哝了一句：“肯定跟零点能有点关系。”姜不愧是老的辣！这句“嘟哝语”给了卡西米尔极大的启示，使他很快通过对零点能的计算，完成了平行导体板这一新情形的独立研究（即得到了我们上面介绍的若干结果）。这一研究因避免了原先依赖的大距离极限，而——如我们很快将会看到的——大大增加了实验验证的可能性。

卡西米尔效应是一种非常微小的量子效应（这并不奇怪，量子效应本身就是以微小著称的）。从上面的公式中可以很容易地计算出有关物理量的数值（当然，一切都是跟平行导体板的间距有关的）。比如能量密度 ρ ——由式（5.4.5）可知——约为

$$\rho = - \frac{4.3 \times 10^{-28}}{d^4} \quad (5.4.8)$$

这里所有的物理量都采用了国际单位制（SI）中的单位，即能量密度 ρ 以焦耳每立方米（J/m³）为单位，平行导体板的间距 d 以米（m）为单位。如果用质能关系式将能量密度折合成质量密度，则约为

$$\rho = - \frac{4.8 \times 10^{-45}}{d^4} \quad (5.4.9)$$

其中质量密度 ρ 以千克每立方米（kg/m³）为单位。这无疑是非常微小的密度。类似地，卡西米尔力的面密度也非常微弱，由式（5.4.7）可知，约为

$$F = - \frac{1.3 \times 10^{-27}}{d^4} \quad (5.4.10)$$

其中面密度 F 以牛顿每平方米 (N/m^2) 为单位。

接下来简单提一下卡西米尔效应的实验确立。前面说过，卡西米尔力的存在为从实验上检验卡西米尔效应提供了直接途径。但是途径虽然有了，测量由式 (5.4.10) 所表示的那么微小的力对实验物理学家来说依然是极大的挑战。唯一有利的，是这种力反比于平行导体板间距的四次方，从而可以通过缩小平行导体板的间距而得到非常显著的“放大”，这堪称实验验证中一个最重要的努力方向（前面提到的因避免大距离极限而大大增加实验验证的可能性，原因正在于此）。比如平行导体板的间距若能缩小到微米 (μm) 量级，力密度就可以增加到千分之一牛顿每平方米的量级。只不过，让两块导体板靠得如此之近而又互不接触绝非易事，通常只有对很小的导体板才能做到。而导体板越小，作用在导体板上的力的总量也将变得越小，从而又反过来增加了检测难度。不过幸运的是，卡西米尔效应并不局限于平行导体板，而早已被理论物理学家们推广到了各种更复杂的情形，比如针对其他形状、非理想导体板、电介质板、有限温度，等等，为实验研究提供了多种选择。那些更复杂的情形由于偏离理想条件而使得理论计算更为复杂，但对于实验验证来说，却因为所要求的条件不那么苛刻，而变得更容易实现。唯一不变的是，无论哪种情形下的卡西米尔效应都是非常微小的。因此，对卡西米尔效应的实验研究虽然从20世纪50年代起就陆续有人在做，却直到90年代后期才开始有了较高精度的验证[\(17\)](#)。

除卡西米尔效应外，还有一些其他量子效应也能在某些特定区域产生负能量。因此，奇异物质的存在是毋庸置疑的，这对于可穿越虫洞来说无疑是“利好消息”。只不过，这个“利好消息”是否真能让我们“获利”，却是大为可疑的。因为所有那些（源于量子效应的）负能量都是非常微小的，而且全都依赖于非常特殊的环境条件的配置才会出现，从而完全不像普通物质那样可以被随心所欲地、独立于其他物体地移动，并建造我们所需要的结构。从这个意义上讲，奇异物质的“奇异”虽是货

真价实的，能否对得住“物质”这一“光荣称号”却大可商榷。起码，它远没有普通物质那样的独立性。奇异物质的存在离不开以特殊方式配置起来的环境条件，它的移动也必须伴随环境条件的变更，就像垂危病人的移动必须伴随着医疗器械的移动一样。而那些无可或缺的环境条件——它们都是由普通物质构成的——完全有可能对奇异物质的“效用”（主要是引力效应）产生重大干扰，甚至彻底地抵消和淹没之[\(18\)](#)。因此，负能量（或奇异物质——如果我们依然愿意这么称呼它的话）虽然确实是存在的，它能否被用来构筑可穿越虫洞，却是一个很大的未知数——而且恐怕是极不容乐观的未知数。

5.5 虫洞的“工程学”

在5.4节中，我们将可穿越虫洞的物质分布会破坏零能量条件这一结论具体化了，那就是构筑可穿越虫洞需要用到所谓的奇异物质（或负能量物质）。并且以卡西米尔效应为例，对奇异物质的产生方式做了介绍。我们看到，奇异物质由于是靠量子效应产生的，其数量往往是非常微小的。卡西米尔效应从理论上被提出到实验上被验证，相隔了差不多半个世纪的时间，主要原因也正在于此。这种数量上的微小对于构筑可穿越虫洞来说当然是坏消息，但究竟坏到什么程度呢？还得看“供求关系”的另一边——即构筑可穿越虫洞究竟需要多少奇异物质。如果需要量很少，情况也许就不算太坏；反之，则不容乐观。

那么，构筑可穿越虫洞究竟需要多少奇异物质呢？在本节中，我们将转入虫洞研究的所谓“工程学”方面，来对这一问题作些探讨。

很明显，构筑不同结构的虫洞所需要的奇异物质的数量可能是不同的，因此这一问题的答案取决于虫洞的具体结构。不过，涉及虫洞具体结构的计算通常是不太容易的，而且那种不太容易的计算往往还没什么特别的重要性（因为往往没什么理由认为某种结构比另一种结构更值得计算），从而给人一种吃力不讨好的感觉。怎么办呢？物理学家们想到了一个聪明的点子，那就是近似——既然选不出最值得计算的虫洞结构，不如只做近似计算，起码还能图个简单。在广义相对论研究中，有一种很常用的近似叫做“薄层”（thin shell）近似，它的要点是将物质分布近似为薄层分布，虫洞物理学上的许多计算就采用了这一近似。在数学上，所谓的薄层是被假定为无限薄的（即厚度趋于零），相应的物质分布呈现 δ 函数的形式。

薄层近似不仅具有数学上的简单性，从“工程学”的角度讲也不无优点。因为尽管如我们在5.4节末尾所说，奇异物质能否被用来构筑可穿

越虫洞是一个“极不容乐观的未知数”，但假如它能够，那么从“工程学”的角度讲，很自然的做法就是将这种数量非常微小的奇异物质的用量尽可能减少。而减少用量的途径之一就是将它的使用范围尽可能局限起来，比如局限在薄层上。除此之外，薄层近似还有一个很大的优点，那就是它在数学上的简单性使人们可以在其他方面适当地引进复杂性，比如引进非球对称性。我们在后文中将会看到，这对于更全面地分析虫洞的可行性是有很大大益处的。

利用薄层近似，物理学家们对构筑可穿越虫洞所需要的奇异物质的数量进行了计算。计算的结果虽仍与虫洞的具体结构——即薄层形状——有关，但托薄层近似的福，可以用一个系统的公式来概括了。在这里，我们将再次采取“中庸之道”，用量纲分析法来推出那个系统的公式。我们用 M 表示构筑可穿越虫洞所需要的奇异物质的数量。很明显， M 与虫洞喉咙的大小——也称为虫洞的半径—— r_0 有关。除此之外，作为广义相对论的结果，它当然也跟万有引力常数 G 及光速 c 有关，因此 $M = f(r_0, G, c)$ 。与上节中推导卡西米尔效应的做法完全类似，我们可以通过对量纲进行比较，得到唯一的函数形式（感兴趣的读者不妨自己推导一下），即

$$M \propto -\frac{c^2}{G}r_0 \quad (5.5.1)$$

这里，我们将负号单独列出，以表示奇异物质的能量为负这一重要特征。考虑到——如前所述——计算结果与薄层形状有关，因此完整的结果还需包含一个与薄层形状有关的因子 F_s ，即

$$M = -\frac{c^2}{G}F_s r_0 \quad (5.5.2)$$

其中比例系数已被吸收进了 F_s 之中，因此比例符号“ \propto ”变成了等号“ $=$ ”。 F_s 作为与可穿越虫洞的具体结构有关的因子，承载了计算中的

细节，但在数量级意义上对结果没有显著影响。式（5.5.2）对于更一般的虫洞也基本成立，只不过 F_s 的计算通常要复杂得多。

将万有引力常数 G 及光速 c 的数值代入，不难得到与上述公式相对应的数量级意义上的数值结果（感兴趣的读者可以核验一下）

$$M = -M_{\odot} r_0 \quad (5.5.3)$$

其中 M_{\odot} 为太阳的质量（数量级为 10^{30}kg ），虫洞的半径 r_0 以千米（km）为单位。

把这一结果与上节所得到的有关奇异物质质量密度的数值结果式（5.4.9）相比较，我们看到，构筑可穿越虫洞在“工程学”上绝对是一个空前巨大的挑战。因为太阳作为一个天体，它的质量乃是所谓的“天文数字”，而奇异物质来自量子效应，其数量是典型的“微乎其微”。因此，构筑可穿越虫洞乃是要将“微乎其微”的量子效应累积成“天文数字”般的巨大数量，其难度是可想而知的（更不用说这种量子效应能否被累积还是一个未知数）。

更糟糕的是，难度上的挑战并未到此为止。因为依靠相当于一个太阳质量的奇异物质所能构筑的，乃是一个半径约1000米的虫洞。而接下来的问题是：那样的虫洞真的可穿越吗？

初看起来，1000米算是不小的半径，假如虫洞类似于普通隧道的话，那样的半径应该足以让大小可观的星际飞船穿越了。看过科幻影片的人想必都对星际飞船穿越虫洞的特技处理留有深刻印象。从影片中看，星际飞船的周围充斥着星光组成的绚丽幻象，它所穿越的则似乎只是一条狭小的通道（见图5.2）。但实际上，星际飞船穿越虫洞的情形远比科幻影片所设想的来得复杂。为了让飞船及乘员安全穿越虫洞，几何半径的大小并不是最主要的问题。那么什么才是最主要的问题呢？是“活着”——也就是5.3节列出的可穿越虫洞所需满足的条件之第（2）条：穿越过程中遇到的应力是人体能够承受的。



图5.2 星际飞船穿越虫洞的特技处理

下面我们就以球对称虫洞为例，来谈谈这一条件。具体地说，穿越虫洞时，星际飞船及乘员将会遇到两种不同类型的应力：一种来自虫洞物质本身的张力，另一种则是虫洞引力场所产生的潮汐力。

我们先看前者。在5.3节中，我们曾经得到过一个当时称之为“对于探讨虫洞是否真的‘可穿越’有着重要影响”的公式，即式（5.3.5），它给出的是球对称可穿越虫洞喉咙处的物质张力。考虑到虫洞解的连续性，该公式也近似地给出了喉咙附近一个小区域内的物质张力。由于球对称可穿越虫洞的喉咙附近是零能量条件遭到破坏的区域（参阅5.3节），从而也就是奇异物质的分布区域（参阅5.4节），因此该公式给出的张力也被称为是奇异物质的张力⁽¹⁹⁾。

我们为什么将这个张力称为“对于探讨虫洞是否真的‘可穿越’有着重要影响”呢？关键在于它的数值。这数值从式（5.3.5）中是难以直接看出的，因为物理常数被略去了。现在我们就将之恢复起来。恢复的方法大家应该已经很熟悉了，就是量纲分析法。恢复的结果是（请读者自行验证）

$$\tau = \frac{c^4}{8\pi G r_0^2} \quad (5.5.4)$$

这里我们略去了等式左端表示喉咙位置的 r_0 ，以及等式右端来自约定，

从而并无本质意义的负号。将物理常数的数值代入便可得到数值结果为

$$\tau = \frac{5 \times 10^{10}}{r_0^2} \quad (5.5.5)$$

其中张力 τ 以牛顿每平方米 (N/m^2)，或等价地以焦耳每三次方米 (J/m^3) 为单位，虫洞半径 r_0 以光年 (light year) 为单位。

这里需要特别注意的是，虫洞半径所采用的不是像“米”或“千米”之类的日常单位，而是光年这样的巨大单位。因此，上式表明一个半径为一光年的球对称虫洞喉咙附近的张力大小约为 $5 \times 10^{10} \text{N/m}^2$ 。那么 $5 \times 10^{10} \text{N/m}^2$ 又是一个什么概念呢？它相当于在每平方米的面积上压上500万吨的重物，或大致相当于物质在原子线度上所能承受的张力[\[20\]](#)。这意味着半径为一光年的球对称虫洞喉咙附近的张力如果作用在物质上，将足以破坏原子结构[\[21\]](#)。由于物质的宏观结构——以所能承受的张力而论——要比原子结构脆弱得多，因此，无论星际飞船还是乘员，在那样的张力作用下都将是“不堪一击”的。而倘若虫洞的半径不是一光年，而是更小，则喉咙附近的张力将会更大（因为张力的大小与虫洞半径的平方成反比）。倘若虫洞的半径只有1000米，则张力将高达不可思议的 $5 \times 10^{36} \text{N/m}^2$ ，相当于每平方米的面积上压上5亿亿亿吨的重物，这无论从“工程学”的意义上，还是从已知物质的性质上讲都是近乎荒谬的。

由此我们看到，别说是半径1000米的球对称虫洞，就连半径一光年的球对称虫洞喉咙附近的张力如果作用在物质上，也并非星际飞船或人体所能承受的。那我们还能依靠什么呢？就只能靠式(5.5.5)所给出的张力与虫洞半径的平方成反比这一特点，通过半径比一光年更大的虫洞来进行星际旅行了。但这在“工程学”上同样是近乎荒谬的。因为式

(5.5.3)表明构筑可穿越虫洞所需的奇异物质的数量正比于虫洞半径，而且构筑半径为1000米的虫洞所需的奇异物质的数量就相当于一个太阳

的质量。由于一光年约为十万亿千米，因此构筑一个半径为一光年的虫洞所需的奇异物质的数量约相当于太阳质量的十万亿倍！构筑能让星际飞船及乘员平安穿越的真正意义上的球对称可穿越虫洞所需的奇异物质的数量甚至比这更多。如此数量的物质（比整个银河系的质量还大得多）别说是奇异物质，就算是普通物质也实在是惊世骇俗的。

在索恩应萨根的电话咨询而研究虫洞物理学的时候，他曾经用一个特殊的视角来判断虫洞作为星际旅行通道的可行性，那就是：什么样的事情是物理定律允许无限发达的文明（infinitely advanced civilization）去做的？之所以采用这样的视角，是因为作为理论物理学家，他所关注的仅仅是物理定律是否允许可穿越虫洞存在，而并不在意具体的实施难度——他把后者扔给了“无限发达的文明”去操心，因为“无限发达的文明”按定义就是能做到物理定律所许可的一切事情的。他并且把由这一特殊视角所提出的问题统称为“萨根式问题”（Sagan-type question）[\[22\]](#)。不过，即便采用“萨根式问题”的特殊视角，上面提到的构筑球对称可穿越虫洞所需要的奇异物质的数量仍是一个灰色地带，甚至可以说是对“萨根式问题”本身的一种诘难。因为单纯从物理定律上讲，那样的可穿越虫洞或许是可以存在的，但实际上却几乎可以断定，操控数量如此之巨的奇异物质对于生活在可观测宇宙中的任何文明——哪怕是“无限发达的文明”——都是不太可能的。因此，上述可穿越虫洞可以说是介于理论上不可能与实际上不可能之间。而萨根所设想的“纵横交错的虫洞”用上述可穿越虫洞来衡量，则是天方夜谭。

接下来再看看星际飞船及乘员将会遇到另一种应力：虫洞引力场所产生的潮汐力。潮汐力的根源是引力场的不均匀性。由于这种不均匀性，物体上两个相距 Δx 的点所感受到的引力加速度会有所不同。我们知道，引力加速度本身的效应是可以通过等效原理消去的，但这种不同点所感受到的引力加速度 a 的不同（即 Δa ）是无法消去的，从而将体现为一种额外的作用，那就是潮汐力。潮汐力是一种常见的力，因为它在地

球上就有一种明显的表现形式：潮汐（其名称也由此而来）。潮汐力的计算是广义相对论中的标准内容，其加速度的表达式为

$$(\Delta a)^\mu = -R^\mu_{\alpha\nu\beta} V^\alpha (\Delta x)^\nu V^\beta \quad (5.5.6)$$

其中 R 是引力场的曲率张量， V 是物体的四维速度。对于球对称虫洞，上述加速度可以分解为径向部分 $(\Delta a)_\parallel$ 与横向部分 $(\Delta a)_\perp$ ，结果分别为（感兴趣的读者可以用度规式（5.3.1）具体算一下）

$$(\Delta a)_\parallel = \left\{ \left(1 - \frac{b}{r} \right) \left[-\varphi'' - (\varphi')^2 \right] + \frac{(b'r - b)\varphi'}{2r^2} \right\} (\Delta x)_\parallel \quad (5.5.7)$$

$$(\Delta a)_\perp = \frac{\gamma^2}{r^2} \left[(r - b)\varphi' + \frac{v^2}{2} \left(b' - \frac{b}{r} \right) \right] (\Delta x)_\perp \quad (5.5.8)$$

其中 v 是物体的三维速度值， $\gamma = (1 - v^2)^{-1/2}$ 是洛伦兹因子。

有了这些结果，原则上就可以计算出星际飞船及乘员穿越球对称虫洞时将会遇到的潮汐力了。但是与计算奇异物质数量的情形相似，结果显然也是跟球对称虫洞的具体结构有关的（此外还可能跟星际飞船的运动方式有关——因为跟飞船运动有关的 v 和 γ 也出现在了公式中）。我们采用的点子也类似，那就是近似。不过，原先用过的薄层近似在这里是不恰当的，因为它所具有的 δ 函数形式的物质分布反映到潮汐力上会产生奇异性。那么，什么样的近似适合潮汐力的计算呢？最简单的一种是所谓的近施瓦西（proximal Schwarzschild）近似。这是对施瓦西度规的一个貌似细微的变更——即变更为

$$ds^2 = \left(1 - \frac{2m}{r} + \frac{\epsilon}{r^2} \right) dt^2 - \left(1 - \frac{2m}{r} \right)^{-1} dr^2 - r^2 d\Omega^2 \quad (5.5.9)$$

其中 $\epsilon > 0$ 是一个很小的参数^{[\[23\]](#)}。很明显，除了 $r = 2M$ 附近的一个小区域外，近施瓦西度规处处都很接近施瓦西度规。不过， ϵ 从数值上讲虽然细微（即所谓“貌似细微”），对时空的整体性质却有着非同小可的影

响，因为引进了 ϵ 之后， $r=2M$ 处的 g_{tt} 就不再为零，从而使得原本出现在 $r=2M$ 处的视界不复存在了。这就克服了将爱因斯坦-罗森桥视为星际旅行通道所遭遇的本质上是来自视界的困难。我们在5.2节中曾经说过，爱因斯坦-罗森桥所具有的连接两个渐近平直时空的特性，起码在表观上与虫洞有着异曲同工之处。来自视界的困难一旦被克服，这种“异曲同工”性就可以由表观转变为实质，爱因斯坦-罗森桥也就荣升为虫洞了。这种虫洞被称为近施瓦西虫洞，它的喉咙位于 $r=2M$ 处⁽²⁴⁾。

由于近施瓦西虫洞的度规在远离喉咙处与施瓦西度规十分接近，因此潮汐力也就基本上等于施瓦西度规中的潮汐力，其加速度的表达式为（这也是广义相对论中的标准结果，感兴趣的读者可通过式（5.5.7）和式（5.5.8）进行验证）

$$(\Delta a)_{\parallel} = \frac{2GM}{r^3}(\Delta x)_{\parallel} \quad (5.5.10)$$

$$(\Delta a)_{\perp} = -\frac{GM}{r^3}(\Delta x)_{\perp} \quad (5.5.11)$$

其中横向加速度 $(\Delta a)_{\perp}$ 的表达式带有一个负号，表示物体在横向上会被压缩；而径向加速度 $(\Delta a)_{\parallel}$ 的表达式不带负号，表示物体在径向上会被拉伸。我们前面提到过，半径为1000米的虫洞喉咙附近的张力会达到不可思议的数值，从而无论从“工程学”的意义上，还是从已知物质的性质上讲都是近乎荒谬的。现在我们可以以近施瓦西虫洞为例，从潮汐力的角度来检视一下半径为1000米的虫洞。因为潮汐力的作用，当星际飞船接近虫洞时，飞船上的乘员会渐渐感到自己的身体在沿虫洞的方向（即径向）上被拉伸，而在与之垂直的方向（即横向）上被挤压。这种感觉在一开始只是让人稍有些不适而已。但由于潮汐力的大小反比于距离的三次方，因此随着飞船的继续前行，潮汐力会迅速增加，距离每缩小到十分之一，潮汐力就会增加至1000倍。当飞船距离虫洞中心还有

1000千米的时候，潮汐力的大小就基本达到了人体所能承受的极限。如果这时候飞船还不立刻折回的话，所有乘员都将在致命的潮汐力作用下丧命。再往前飞行一段距离，则飞船本身也将在潮汐力的作用下解体。最终从虫洞另一端飞出的将不是意气风发的星际旅行家，而是飞船和人体的残骸⁽²⁵⁾！

这就是试图穿越半径1000米的近施瓦西虫洞的星际旅行家的下场。因此，从潮汐力的角度看，半径1000米的虫洞——起码就近施瓦西虫洞这一例子而言——是无法作为星际旅行的通道的。不仅如此，包含潮汐效应在内的虫洞的巨大引力场会在周围一个很大的范围内产生毁灭性的破坏力，科幻小说或电影中的那些出入口位于行星表面的虫洞从这个意义上讲，都是极不可能的。在萨根的故事中，曾有人反对将外星智慧生物传来的蓝图付诸实施，因为他们担心那可能是一个能够毁灭地球的装置，他们的担心是很有道理的。那么，什么样的虫洞产生的潮汐力才是在整个穿越过程中都是人体能够承受的呢？对近施瓦西虫洞来说，大约要半径在几万千米以上才行（读者可以自行估算一下）。这与来自张力的条件（半径一光年以上）相比虽然宽松得多，依然是非常大的尺度（相应地，所需要的奇异物质的数量则相当于太阳质量的几万倍以上）。

以上就是对虫洞的所谓“工程学”讨论，就讨论中所涉及例子而言，基本结论是：可穿越虫洞的半径必须很大，才能保障穿越过程中遇到的应力是人体能够承受的。相应的，所需奇异物质的数量也必须很大——大到几乎没有实现的可能。不过，这一结论并非无懈可击，因为讨论中涉及例子毕竟是有限的。那么，在更一般的情形下是否有可能绕过这一结论呢？对此我们在后文中还会做进一步讨论。

5.6 由虫洞到时间机器

关于虫洞物理学，还有一个十分有趣的衍生课题，是我们要在本节中讨论的，那就是虫洞与时间机器的关系。这种关系听起来有些玄妙，对于熟悉相对论的人来说其实是不难想到的，因为虫洞作为连接两个时空区域的捷径，可以说是在某种意义上实现了超光速旅行。或者换一个说法，是在某种意义上使得类空事件之间的物理联系（因果联系）成为了可能^{[\(26\)](#)}。另一方面，众所周知的是：在相对论中，类空事件的时序是与参照系的选择有关的。因此虫洞从某种意义上讲可以说是开启了因果颠倒的大门，而这正是时间机器的重要特征。因此，虫洞与时间机器是有可能存在联系的。

不过，在索恩和莫里斯研究虫洞时，也许是因为对虫洞本身倾注了太多注意力的缘故，忽视了像虫洞与时间机器的关系那样的衍生课题。但幸运的是，那样的忽视并没有持续太久。1986年底，索恩和莫里斯在芝加哥参加了一次天体物理学会议，身为老板的索恩做了一个学术报告（但不是有关虫洞的），当时还是研究生的莫里斯则四处与人聊天，混个脸熟，其中有位聊天对象是中康涅狄格州立大学（Central Connecticut State University）的物理学家，名叫罗曼（Thomas Roman）。索恩和莫里斯有关虫洞的研究当时已接近完成，莫里斯就在聊天中介绍了该项研究。正所谓“当局者迷，旁观者清”，罗曼一听完莫里斯的介绍马上就问道：如果虫洞能让人以比光速更快的速度跨越空间距离，那是否意味着能用它旅行到过去？“一语点醒梦中人”，罗曼的问题让索恩和莫里斯大为震动，以至于索恩后来述及此事时再次“沉痛”检讨了自己的愚蠢^{[\(27\)](#)}。

罗曼的问题使索恩和莫里斯立刻投入到了对虫洞与时间机器关系的研究之中（值得一提的是：学术交流是互惠的，受聊天的影响，罗曼自

已后来也涉足了虫洞与时间机器领域的研究），并在论文中添加了一个新的部分，叙述了一种利用两个虫洞构筑时间机器的方案。不久之后，他们又提出了一个新方案，将两个虫洞减少为一个。那个新方案就是目前文献中用虫洞构筑时间机器的标准方案（当然，不同文献在具体表述上各有特色，但基本思路是一致的）。

下面我们就对这个标准方案作一个简单介绍。这个方案是通过三个步骤构筑时间机器的：

（1）构筑一个虫洞（或利用一个现成的虫洞）。在介绍时间机器的构筑时，为简单起见，人们通常略去虫洞本身的“工程学”细节，而将之抽象为一个对相距 L 的两条世界线 $(t, 0, 0, 0)$ 和 $(t, 0, 0, L)$ 作等时连接的通道^{[\[28\]](#)}。

（2）设法在虫洞的入口和出口之间产生一个时间差。这有许多办法可以做到。比如可以通过让出口相对于入口作高速运动后返回，利用狭义相对论的时钟延缓效应来做到^{[\[29\]](#)}，或通过将出口置于与入口不同的外部引力场中，利用引力场的时钟延缓效应来做到，等等。假定所产生的时间差为 T ，则这一步完成时虫洞所连接的两条世界线将成为 $(t, 0, 0, 0)$ 和 $(t+T, 0, 0, L)$ 。

（3）将虫洞出入口之间的距离 L 缓慢地缩小，直至接近于零^{[\[30\]](#)}。这一步完成后，虫洞所连接的两条世界线将变成 $(t, 0, 0, 0)$ 和 $(t+T, 0, 0, 0)$ 。

由上述三个步骤所构筑的连接 $(t, 0, 0, 0)$ 和 $(t+T, 0, 0, 0)$ 的虫洞就是一种时间机器，因为由入口往出口方向穿越这一虫洞会跨越幅度为 T 的时间，由出口往入口方向穿越这一虫洞则会跨越幅度为 $-T$ 的时间。因此，这一时间机器既可以通向未来，也可以重返过去^{[\[31\]](#)}。

用虫洞构筑时间机器对于科幻作家来说可谓是喜出望外的事情，因为时间机器无论对于科幻作家还是他们的读者，都是吸引力堪与虫洞相比——甚至犹有过之——的东西（谁不想遨游未来或重返过去呢？）。

事实上，早在虫洞、黑洞甚至相对论问世之前，科幻作家们就已创作出了不少有关时间机器或时间旅行的故事。比如著名英国科幻作家威尔斯（H. G. Wells, 1866—1946）早在1895年（即比狭义相对论的问世还早了十年）就发表了名为《时间机器》（The Time Machine）的著名小说。用虫洞构筑时间机器的设想虽问世很晚，但由于是从科学土壤里萌发出来的，且与虫洞这一热门概念紧密相通，因此大有后来居上的势头。在新的科幻小说或影片中，像普通机器那样“老土”的时间机器已基本淡出，取而代之的往往是用类似于穿越虫洞的文字或镜头表示的时间旅行^[32]。

但物理学家们的态度要谨慎得多，因为时间机器在物理学上一向是烫手的山芋。其中最烫手的就是由此引发的几乎无穷无尽的因果悖论^[33]。那些悖论是如此棘手，以至于在很长的时间里，时间机器被多数物理学家视为是幻想领域的东西，而非严肃课题。比如索恩研究时间机器的消息传出后，就曾接到过朋友的电话垂询，担心他会“走火入魔”。甚至他本人也担心虽然自己“老了，无所谓”，但此类消息有可能会影响学生的前程。当然，经过若干位物理学家的共同努力，如今的情况已有所改变，但时间机器依然是一个几乎在所有方面都有着高度争议性的话题。索恩自己的态度也非常谨慎。我们知道，索恩曾经跟霍金就宇宙监督假设打过赌，其中霍金的立场是“上帝憎恶裸奇点”。霍金对时间机器有一个类似的立场，叫做“自然憎恶时间机器”（Nature abhors a time machine）。用学术术语来说，霍金的立场叫做“时序保护假设”（我们在2.6节末尾提到过这一假设及其依据），它通俗地说就是认为时间机器起码在宏观尺度上是不存在的。对于霍金的这一立场，索恩也曾考虑过打赌，但思虑再三后，却未敢下注。其原因用他自己的话说，是“我喜欢跟霍金打赌，但只打我有合理可能性会赢的赌”。索恩并且承认，就时序保护假设而言，无论从直觉还是亲自做过的计算来看，他都觉得“霍金有可能是对的”。

如果霍金是对的，即“自然憎恶时间机器”，那么用虫洞构筑时间机器就应该是不可能的。但索恩——如上面三个步骤所显示的——又明明已经至少在理论上找到了构筑时间机器的方法，这该做何解释呢？对此，美国物理学家沃尔德和杰罗奇提出了一种可能性，那就是虫洞一旦被构筑成时间机器，会在构筑成功的一瞬间就自动毁灭。因为在那一瞬间，任何细微的辐射都可以通过虫洞返回过去，与它自身相叠加。这种叠加过程可以在零时间内重复无穷多次，从而产生毁灭性的自激效应 [\(34\)](#)。

除去理论层面的考虑外，用虫洞构筑时间机器还面临着实际层面上的诸多困难。事实上，构筑可穿越虫洞本身就已如此困难，通过种种操作在其出入口之间产生可观的时间差无疑更是难上加难。从这个角度看，用虫洞构筑时间机器的前景是相当渺茫的。

5.7 讨论

我们有关虫洞物理学的介绍到这里就大体结束了。依照前面各专题所遵循的结构惯例，这最后一节将做一些略带延伸性的讨论。

首先要讨论的话题是：虫洞究竟有没有可能被构筑出来？读者也许会觉得奇怪？我们不是一直就在讨论这个问题，并且还得出了诸如构筑可穿越虫洞需要奇异物质之类的具体结论吗？怎么又绕回去了呢？不是绕回去，而是不同的问题。因为我们前面讨论的其实只是如何维持一个虫洞，而不是如何构筑一个虫洞，后者是指从无到有地产生一个虫洞，因而是完全不同的问题。

回到5.1节用过的苹果的比喻上来：如果人类生活在苹果的表面上，而虫洞是穿越苹果内部的通道，那么很明显，那样的通道是对苹果表面拓扑结构的一种改变。因此，从无到有地产生虫洞意味着“拓扑改变”（topology change）。关于这种拓扑改变的含义，物理学家们曾有过分歧。虫洞概念的“开山祖师”惠勒思考这一概念时，所针对的乃是时空，即认为时空的拓扑结构有可能因量子涨落而改变（参阅5.2节）。这个想法曾被一些物理学家所继承，但也有一些物理学家表示了怀疑，因为稍稍细想一下，就会发现这个想法是有问题的。所谓时空，顾名思义，乃是时间与空间的合称，是已经包含了时间的。既然如此，时空作为一个四维流形，按定义就是完全确定的，它的任何性质都不可能发生变化——因为它之外并不存在可用来定义“变化”的时间。

那么，拓扑改变指的究竟是什么呢？比较合理的看法是将时空流形分解为空间部分和时间部分，然后将拓扑改变定义为空间部分的拓扑相对于时间的变化^[35]。

那么，如此定义的拓扑改变有可能出现吗？物理学家们已经进行了一些研究，并得到了一些结果。比如1967年，杰罗奇（又是此人！）证

明了这样一个结果：如果两个紧致的类空超曲面 Σ_1 和 Σ_2 能通过一个紧致的时空 M 相衔接，并且在 M 上能定义明确的时间方向，且不包含闭合类时曲线，则 Σ_1 和 Σ_2 必定有相同的拓扑结构。霍金及同事也证明了一系列类似的结果。用通俗的话说，这些结果意味着在有界（“紧致”）的且具有良好的因果性质（“能定义明确的时间方向，且不包含闭合类时曲线”）的时空中，拓扑改变是不可能的。除此之外，如果假定时空是全局双曲的，则还可以证明这样的结果：全局双曲时空必定可以分解为空间部分和时间部分，其中空间部分（即对应于任一时间坐标的类空超曲面）全都具有相同的拓扑结构。这同样意味着拓扑改变是不可能的（而且顺带保证了定义拓扑改变所需要的“将时空流形分解为空间部分和时间部分”是可以做到的）。

虽然这些结果所要求的条件不无苛刻性，且大都极难验证，但它们的存在无疑为构筑虫洞的可能性投下了巨大的阴影。这可以说是在我们前面介绍过的有关可穿越虫洞的诸多困难之外又增添了一种新困难，而且有可能是原则上不可克服的困难⁽³⁶⁾。不过，这也并非是给虫洞物理学判处了死刑。因为即便这些结果成立，也还存在这样的可能性，那就是空间的初始结构已经包含虫洞。如果这样的话，我们（或“无限发达的文明”）所要做的只是把已经存在的虫洞“做大”，使之成为可穿越虫洞。而这——如我们在前面各节中所讨论的——虽然极端困难，却不是原则上不可能的。

关于虫洞的另一个讨论话题是，虫洞究竟是不是星际旅行的捷径？很多人——包括科普作家乃至物理学家——提到虫洞时都会想当然地认为那是星际旅行的捷径，比如像图5.3所示的那样。其实，虫洞并不一定是连接两个时空区域的捷径。图5.3本身就是对这一点很好的诠释，因为从图中可以看到，虫洞之所以成为捷径，完全是因为外部空间被弯曲成了U形，倘若没有这种堪称离奇的弯曲，虫洞就失去了最大的价值。事实上，即便在苹果这一原始比喻中，虫洞所能起到的捷径作用也

是十分有限的。倘若外部空间本身是接近平直的，则虫洞非但不一定是捷径，反而很可能是绕远。这一点几乎被有关虫洞的科幻、科普乃至专著所一致忽略，对于虫洞的应用来说却是一瓢不容忽视的冷水[\(37\)](#)。

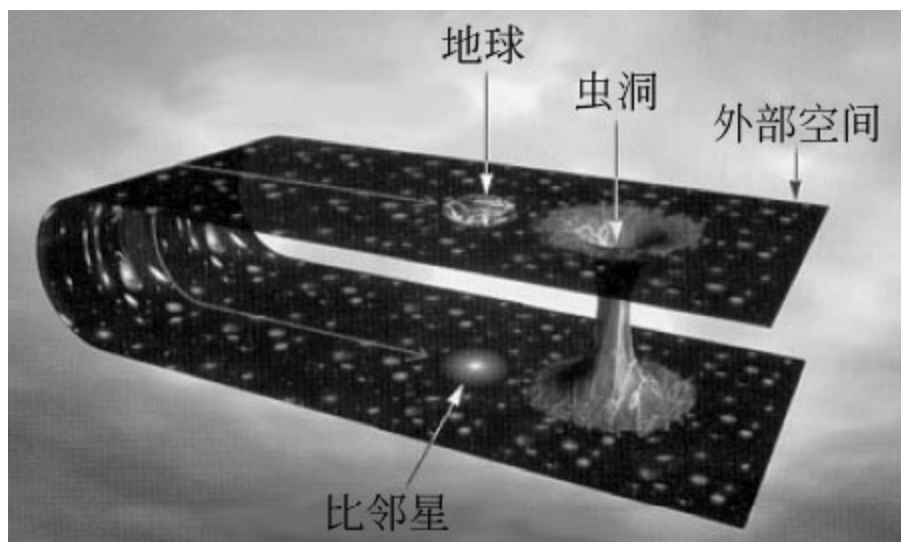


图5.3 典型的虫洞示意图

以上两个讨论话题一个为虫洞的构筑投下了阴影，一个为虫洞的应用泼下了冷水，对虫洞爱好者来说大概都不是喜闻乐见的。接下来讨论一个比较“正面”的话题吧。我们在上节的末尾曾经留下一个问题，那就是有关虫洞“工程学”的讨论及所得到的（很不乐观的）基本结论在更一般的情形下是否有可能绕过？对这个问题物理学家们也已做过一些探讨，得到的结果是勉强可算正面的。

比方说，在薄层近似下（这几乎是唯一能让我们摆脱球对称的近似），有物理学家研究了形状为立方体（或长方体）的虫洞，结果发现，这种虫洞的奇异物质只需分布在立方体（或长方体）的12条棱上即可。这意味着星际飞船可以不必与奇异物质直接接触就穿越虫洞，从而可以绕过最严峻的困难之一，即虫洞物质的张力有可能对星际飞船及乘员造成危害。有意思的是，比这一理论早得多，甚至早在虫洞这一概念因萨根与索恩的交流而在科幻小说中正式登场之前的1968年，著名英国

科幻作家克拉克（Arthur Clarke, 1917—2008）就在名著《2001：太空奥德赛》（2001: A Space Odyssey）中引进了一个他称之为“星之门”（stargate）的概念。这个概念与虫洞的出入口极为相似。不仅如此，克拉克引进的“星之门”的形状恰好也是长方体，与几十年后的虫洞研究形成了有趣的呼应^[38]。

不过，形状为立方体（或长方体）的虫洞从概念上讲虽不存在困难，从“工程学”角度看却绝不容易，因为计算表明，它12条棱上的奇异物质的张力必须达到 10^{43} 牛顿（N）的惊人数值，约相当于普通原子组成的一维结构所能提供的张力大小的一万亿亿亿亿亿亿（ 10^{52} ）倍。因此，这方面的结果虽属正面，却极为勉强，以可行性而论，与不那么正面的结果相比，恐怕只是五十步与一百步之别。

星际飞船及乘员将会遇到的另一种应力——潮汐力——的情况又如何呢？研究表明，那可以通过选择特殊结构的虫洞及限制飞船的速度而加以控制甚至部分消除，从而也是可以绕过的。这一点甚至无须计算就可看出。比如由式（5.5.7）给出的径向潮汐加速度可以通过选择 ϕ 为常数的虫洞而彻底消除。此时的横向潮汐加速度则因为正比于 v^2 ，而可以通过限制飞船的速度加以控制。因此，潮汐力带来的困难是可以绕过的。但是，绕过张力困难（如果可能的话）与绕过潮汐力困难所需满足的条件是不同的，能否同时绕过两者则是一个未知数。

除设法绕过张力困难与潮汐力困难外，关于虫洞还有一种有趣的设想值得一提，那就是将它与所谓的生命传输机（transporter）结合起来。生命传输机是一种将人体分解为基本粒子进行传输，并在目的地予以复现的装置。这种装置在著名美国科幻电视连续剧《星际迷航》

（Star Trek）中有着极频繁的使用，堪称“招牌”技术。由于组成生命的基本粒子本身并无独特性，生命传输机甚至可以简化为只传输复现人体所需的信息。倘若把这种技术与虫洞结合起来，那么通过虫洞传输的将只是组成人体或传递信息的基本粒子。这虽然也绝不是一件容易的事

情，但虫洞“工程学”上最令人头疼的“穿越过程中遇到的应力是人体能够承受的”这一要求，以及由此导致的有可能极为巨大的奇异物质数量，就也许不再是严重问题了，从而将大大增加利用虫洞进行星际旅行的可能性。不过，这种设想也有自己的问题，其中最大的问题是：生命传输机本身的可行性就是一个未知数，而且是很不容乐观的未知数[\(39\)](#)。

在结束本节——从而也结束本专题及全书——之前[\(40\)](#)，还有一条隐伏在前文的线索需要交待一下。我们在5.4节的末尾曾经提到过，奇异物质的存在离不开以特殊方式配置起来的环境条件，我们并且还提到了一种可能性，即那些由普通物质构成的环境条件有可能对奇异物质的“效用”产生重大干扰，甚至彻底地抵消和淹没之。从某种意义上讲，这种可能性已得到了一些新近研究的证实。

20世纪90年代，美国塔夫斯大学（Tufts University）的物理学家福特（Larry Ford）等人对奇异物质进行了研究，结果发现它的负能量密度会无可避免地受到正能量密度的抵消。并且他们证明了一个定量的结论，即负能量密度必须满足一个被称为“量子不等式”（quantum inequality）的约束条件。用数学公式来表示，所谓量子不等式是这样

一个不等式：

$$\int \rho f(\tau) d\tau \geq - \frac{C}{(\Delta\tau)^4} \quad (5.7.1)$$

其中 ρ 是负能量密度； $f(\tau)$ 是一个对时间自变量 τ 的积分为1，且特征宽度为 $\Delta\tau$ 的“抽样函数”（sampling function）； C 是一个常数（通常小于或远小于1）。这个不等式的物理意义是：负能量密度的数值越大，所能维持的时间就越短。

量子不等式有很大的普遍性，不仅适用于平直空间，也适用于弯曲空间，它的发现引起了不少物理学家的兴趣。这是因为我们在本书的前

面各专题——即奇点与奇点定理、正质量定理、宇宙监督假设——中介绍过的各种经典广义相对论的结果都依赖于能量条件，从而都会受到负能量密度的威胁。而量子不等式的存在似乎为这种威胁设下了限制。这种限制能否在一定程度上挽救那些经典结果？这是不少物理学家感兴趣的问题。不过，与负能量密度对那些其他经典结果构成威胁相反，虫洞物理学却偏偏是需要负能量密度的，因此为负能量密度设下限制的量子不等式对其他经典结果或许是福音，对虫洞物理学却很可能是凶兆。

这种凶兆究竟“凶”到什么程度呢？福特等人针对球对称虫洞的情形进行了估算，结果表明，在一般情况下，虫洞只能是微观的，其半径将会有个约为 10^{-31} 米（相当于原子线度的十万亿亿分之一）的上限。那样的虫洞当然不仅绝不是任何星际飞船能够穿越的，就连与生命传输机相结合都会是很成问题的。不过，福特等人也并未将所有的希望都掐灭，在他们分析过的情形中，原则上也有结构非常特殊的大型虫洞可以存在，但要求是：奇异物质必须分布在厚度比原子核还小——甚至小得多——的薄层内。这样的分布在理论上是不陌生的（几乎就是薄层近似），但实际上如何实现却是彻底的未知数。其他研究者也做了种种估算，得到的结果总体上也是不乐观的，甚至互有争议。唯一可以确定的是，迄今尚不存在一个哪怕只是理论上有希望的虫洞解，能解决前面提到过的所有困难。因此，在虫洞这个课题上，幻想与现实之间的距离还是相当遥远的。

本书到这里就完全结束了。最后要提醒读者的是，本书所介绍的各个专题都还在继续发展之中。很多细节可能还有推敲余地，很多结果可能还有改进空间，很多近似还有待突破，很多线索还有待探讨。所有这些都是让物理学家们每天清晨满怀兴趣 and 希望走进办公室的动力之一，也是值得年轻人思考和参与的课题。广义相对论是一个优美的理论，优美得足以让人产生再无研究可做的错觉，如果本书能使某些读者注意到这个优美理论的某些尚有可为的前沿课题，甚至产生兴趣，那作

者的努力就得到了最好的回报。

注释

[\[1\]](#) 惠勒不仅是索恩的导师，还与索恩及美国物理学家米斯纳（惠勒的另一位学生）合作撰写了有广义相对论“圣经”之称的著名教材《引力》（Gravitation）。

[\[2\]](#) 因为这一原因，这一由施瓦西度规变形而来的爱因斯坦-罗森桥有时也被冠以虫洞之名，称为施瓦西虫洞（Schwarzschild wormhole）。

[\[3\]](#) 这一点常常被过分强调，但其实潮汐力的大小是随黑洞质量的增加而减小的，因此只要黑洞质量足够大，潮汐力是可以变得足够小的。比如对于施瓦西黑洞来说，只要质量在几万个太阳质量以上，视界附近的潮汐力就能减小到人体所能承受的程度。这种质量虽绝非单个恒星的引力坍塌所能达到，对于据信普遍存在于星系中心的巨型黑洞来说却是小菜一碟。因此，潮汐力作为一种困难，并没有通常渲染的那样严峻。当然，许多科幻小说中的星际旅行通道的出入口往往建在行星上，或行星附近，那样的设想与上述巨型黑洞倒确实是完全不相容的。

[\[4\]](#) 当然，科幻作品中还有其他一些有关星际旅行的设想，比如科幻作家阿西莫夫（Isaac Asimov, 1920—1992）在著名的银河帝国系列（Galactic Empire series）和基地系列（Foundation series）等作品中就频繁采用了所谓“跃迁”的方式来进行星际旅行。不过这种“跃迁”背后的理论基础近于空白。另外还有一种很流行的设想是所谓的曲速引擎（warp drive），限于篇幅我们不作介绍了。

[\[5\]](#) “普朗克长度”是由万有引力常数（gravitational constant）、普朗克常数（Planck's constant）及光速（speed of light）构成的具有长度量纲的量，数量级为 10^{-35}m 。

[\[6\]](#) 这里还有一点可以补充（或作为这一理解的应用），那就是可穿越虫洞的出入口所连接的渐近平直时空既可以位于同一个宇宙中，也可以位于不同宇宙中，前者被称为“宇宙内”（intra-universe）可穿越虫洞，后者被称为“宇宙间”（inter-universe）可穿越虫洞。这两种可穿越虫洞的主要区别在于时空的大尺度拓扑结构，而虫洞本身的结构可以视为相同，因此在我们的讨论中将不予区分。

[\[7\]](#) 我们对条件的罗列与索恩和莫里斯的原始论文有所不同，但实质是一致的。

[\[8\]](#) 依据这一物理意义， $b_+(\infty)$ 和 $b_-(\infty)$ 的有限性可视为条件（5）（物质的数量是

可观测宇宙可以提供的)的要求。

[\[9\]](#) 粗看起来, $e^{-2\varphi}(1-b/r)$ 在“喉咙”处为零是式 (5.3.2) 的推论。不过, 它其实还有赖于前面提到的另一个结果, 请读者想一想是什么结果。

[\[10\]](#) 忘记了什么是“全局双曲时空”的读者请参阅2.4节。

[\[11\]](#) “奇异物质”这一术语有时还被用来表示其他含义, 比如由所谓“奇异重子”(exotic baryon)组成的物质等。在本书中, 我们只用它来表示破坏零能量条件的物质。

[\[12\]](#) 感兴趣的读者可参阅, 比如, 依捷克森(Claude Itzykson)和祖柏尔(Jean-Bernard Zuber)的《量子场论》(Quantum Field Theory)。该书有中文版, 科学出版社1986年出版。

[\[13\]](#) 对物理学中的近似有一定了解的读者都知道, 物理学中所谓的“大”和“小”乃是相对于具体问题的需要而言的, 因此“平行导体板足够大”并不意味着它们看上去一定是庞然大物。在卡西米尔效应的研究中, 平行导体板的“大”乃是相对于间距而言的, 假如间距非常小(比如只有几微米甚至更小), 则看上去很小(比如面积只有几平方厘米)的导体板也可以被认为是“足够大”的。

[\[14\]](#) 确切地说, \hbar 是所谓“约化普朗克常数”(reduced Planck's constant), 它与德国物理学家普朗克(Max Planck, 1858—1947)最早引进的“正牌”普朗克常数 h 的关系是 $\hbar=h/2\pi$ 。不过约化普朗克常数在量子理论中的运用早已达到了“喧宾夺主”的程度, 以至于直接称其为普朗克常数也未尝不可了。

[\[15\]](#) 这一结果无须复杂的额外计算就可得到——只要考虑到洛伦兹协变性要求 T^{ab} 只依赖于 η^{ab} 和 $z^a z^b$ (z 为导体板的法向单位矢量), 以及电磁场的能量动量张量满足 $T^{ab}\eta_{ab}=0$ 这一特殊性质即可。

[\[16\]](#) 由于这一工作, 玻德有时也被视为卡西米尔效应的共同发现者。

[\[17\]](#) 这类验证中最早的一个是华盛顿大学(University of Washington)的物理学家拉莫理奥克斯(Steve Lamoreaux)于1997年发表的(实验完成于1996年), 针对的是一块导体板与一个导体球之间的卡西米尔效应(从而避免了让两块导体板靠得极近而又互不接触的困难), 验证精度约为5%。

[\[18\]](#) 对于平行导体板情形, 事实上可以很容易地证明, 构成环境条件的普通物质(即导体板)的总能量远远超过依赖于它们而存在的负能量的数量。感兴趣的读者不妨利用本节给出

的有关结果，以及关于普通物质的某些“常识”，自行证明一下。

[\[19\]](#) 这里我们虽然强调了“喉咙附近”，但球对称可穿越虫洞的喉咙所在处本身零能量条件也很可能是遭到破坏的，只不过是不能完全排除刚好处于破坏边缘这一特殊情形而已。但这种理论上的边缘情形在实际上出现的概率是可以忽略的。

[\[20\]](#) 物质在原子线度上所能承受的张力可以用很简单的方法来估算，那就是物质在原子线度上的特征能量除以原子的体积，前者约为 10^{-19}J （即电子伏特量级），后者约为 10^{-30}m^3 （即线度约为 0.1nm ），相应的张力约为 10^{11}J/m^3 ，与 $5\times 10^{10}\text{J/m}^3$ 相当接近。

[\[21\]](#) 细心的读者也许注意到了，这里我们用了“如果”一词。这是因为虽然一般说来，当物体穿越一种具有内部张力的结构时，自身也会受到与那种张力同一量级的张力作用，但考虑到我们对奇异物质的性质还所知不多，不能完全排除它与普通物质的相互作用特别弱，从而使穿越它的普通物体所受张力远小于奇异物质内部张力之类的可能性，因此在措辞上特意留有余地。

[\[22\]](#) 不过，虽然索恩将此类问题冠上了萨根的大名，萨根本人却似乎从未直接用过“无限发达的文明”这一概念。他倒是沿袭拓展过苏联天文学家卡达什耶夫（Nikolai Kardashev, 1935—）提出的将星际文明分为 I 类、II 类、III 类的所谓“卡达什耶夫标度”（Kardashev scale）。

[\[23\]](#) ϵ “很小”的具体含义是 $\epsilon \ll M^2$ （感兴趣的读者可以试着恢复这一关系式中的物理常数）。

[\[24\]](#) 感兴趣的读者可以计算一下这一度规所对应的能量动量张量，证实零能量条件会遭到破坏，且喉咙处的张力满足式（5.3.5）。关于近施瓦西虫洞，一个有趣的问题是：由于它的度规与施瓦西度规很相似，而施瓦西度规所描述的天体被认为是普遍存在的，这是否意味着近施瓦西虫洞可以通过对普通天体做适当的“手术”而比较容易地构筑出来？可惜，这个问题尚无答案，因为近施瓦西虫洞的度规与施瓦西度规虽然相似，时空的整体性质却很不相同。在这种情形下，由一种时空变为另一种时空别说难易，就连是否有可能都还是一个未知数（关于后面这一点，后文会有进一步的讨论）。不过，近施瓦西虫洞作为一个容易计算的例子，无疑是很有用的。

[\[25\]](#) 如果把虫洞喉咙附近奇异物质的张力可能起到的破坏作用也考虑进去的话，那么飞船及乘员的残骸还可能被进一步撕裂，由虫洞另一端飞出的可能将是一串无法分辨来源的亚原

子粒子！

[\[26\]](#) 这里所谓的“在某种意义上”是指从外部时空的角度来看。

[\[27\]](#) 读者们想必还记得，索恩的“第一次”检讨是在收到昔日学生佩奇的来信，提醒他可穿越虫洞破坏零能量条件这一结论能通过全局方法得到时所作的（参阅5.3节）。不过从两次检讨本身的时间顺序上讲，其实这次在先，那次在后。

[\[28\]](#) 这里的 L 是虫洞的入口和出口之间沿外部空间（即不通过虫洞）所测得的距离。“等时连接”则是指进入虫洞和离开虫洞的时刻都可以视为 t （这样做的条件是虫洞的长度——从而穿越虫洞所需的时间——相比于 L 可以忽略）。

[\[29\]](#) 这种利用狭义相对论的时钟延缓效应来产生时间差的方式在历史上曾引起过争议，并被冠以“时钟佯谬”之名，但实际上并无佯谬可言。

[\[30\]](#) 这一步并非必需，而且并非一定得让距离 L 接近于零。事实上，距离 L 只要小于时间差 T 就可以了（感兴趣的读者请想一想，这是为什么？另外也请想一想，距离 L “缓慢地缩小”中“缓慢”二字的用意何在？）

[\[31\]](#) 不过，从构造步骤中不难看出，这种时间机器所能让人重返的“过去”不是完全任意的，它必须晚于虫洞被构筑成时间机器的时刻。这一结论——即时间机器不能让人重返它被构筑之前的年代——被认为有可能是普遍成立的。这一结论不仅有很强的“现实”意义，而且还有很重要的理论意义，因为霍金曾经提出过一个针对时间机器的有趣的诘难，那就是：倘若时间机器的构筑是可能的，那为什么我们从未遇到来自未来世界的时间旅行家呢？这个诘难的潜台词是：我们从未遇到来自未来世界的时间旅行家，最有可能的原因是时间机器在整个时间长河中——也就是永远——都不曾被构筑过，而这背后最有可能的原因则是时间机器的构筑是不可能的。但假如时间机器不能让人重返它被构筑之前的年代，霍金的诘难就不攻自破了，我们也就不一定要得出这么悲观的猜测了。

[\[32\]](#) 用虫洞构筑时间机器除了是“虫洞”与“时间机器”这两个热门概念的“强强组合”外，还有一个引人注目的优点，那就是霍金等人在对时间机器进行研究后曾经发现：如果能量密度处处非负，则在任何有限时空区域内建造时间机器的努力都会产生奇点。这一结论被认为是对时间机器的一种严峻挑战。但用虫洞构筑的时间机器却是一个例外，因为虫洞的引进破坏了“能量密度处处非负”这一前提条件。

[\[33\]](#) 这种悖论中最典型的是所谓的“祖父悖论”（grandfather paradox），即时间旅行家重返过去后设法阻止自己的祖父母相识。这个悖论的“悖”理在于：时间旅行家的祖父母若无法相识，他的父亲就不会出生，从而也就不会有时间旅行家自己了。但时间旅行家自己若不存在，又怎能重返过去并阻止祖父母相识呢？

[\[34\]](#) 对于这种可能性，索恩本人持有一定的异议，他认为虫洞喉咙处的奇异物质会因负能量的缘故，而对辐射产生散焦（defocus）作用。这种散焦作用可以显著削弱辐射与它自身相叠加的效果，从而避免时间机器被摧毁。这方面——以及更一般的有关“时序保护假设”——的研究目前仍在进行之中。

[\[35\]](#) 细心的读者也许注意到了，我们此处遇到的问题及解决问题的方法与3.2节所遇到的定义广义相对论动力学的问题及解决的方法是完全相似的。

[\[36\]](#) 读者可能会问：这些理论结果都是经典广义相对论中的结果，量子引力有可能会推翻这些结果吗？这一问题的答案物理学家们也很想知道，可惜迄今为止还没有人能够确知（因为令人满意的量子引力理论本身尚不存在）。不过有迹象表明量子引力理论有可能不会推翻这些结果，比如在量子引力理论候选方案之一的正则量子化方案中，位形空间是由经典广义相对论确定的，从而会自动遵循后者有关拓扑改变的理论结果。

[\[37\]](#) 当然，将外部空间弯曲成U形也许不是原则上不可能办到的事情，从而对于“无限发达的文明”来说也许不算问题。

[\[38\]](#) 不过，克拉克的“星之门”并无理论基础，以长方体作为形状也“纯属巧合”，估计仅仅是为了让它看起来更像“人工”之物。

[\[39\]](#) 关于这方面的介绍，可参阅拙作“生命传输机”（《科学画报》2003年第10期，上海科学技术出版社）。

[\[40\]](#) 细心的读者也许会指出：5.3节列出的可穿越虫洞所需满足的条件中，“在微扰下保持稳定”这一条（条件（6））尚未介绍过。对此我们补述一下。这一条的直接研究是非常困难的，因为无可避免地要面对非球对称的情形。不过，索恩和莫里斯的原始论文对这一条做了一个很有道理的评注，那就是即便虫洞是不稳定的，一个“无限发达的文明”仍可以通过密切监视它的结构，而将不稳定的苗头随时扼杀在摇篮里。因此，可穿越虫洞的“维稳”不存在原则上的困难。

名词索引

A

ADM分解

ADM能量

ADM能量动量

ADM质量

AdS时空

爱因斯坦-罗森桥

爱因斯坦-麦克斯韦体系

爱因斯坦场方程

爱因斯坦张量

B

白洞

标架场

标量场

玻色子

伯克霍夫定理

薄层近似

不完备非类空测地线

C

测地偏离效应

超对称伙伴

超曲面垂直

超引力

潮汐力

虫洞

虫洞物理学

抽样函数

D

大爆炸

狄拉克方程

狄拉克矩阵

F

范德瓦尔斯力

非类空测地线

非时序点集

菲尔茨奖

费米子

封闭陷获面

弗里德曼解

负能量物质

G

高斯-柯达西方程组

共轭点

共形变换

共形紧致化

共形无穷远

孤立体系

广义仿射参数

过去柯西视界

过去柯西展开

H

哈密顿表述

黑洞

黑洞唯一性定理

黑洞无毛发定理

霍金-彭罗斯奇点定理

J

奇点

奇点定理

迹能量条件

几何动力学

渐近平直时空

近施瓦西虫洞

K

卡西米尔力

卡西米尔效应

柯西面

柯西展开

壳层穿越奇点

壳层会聚奇点

可穿越虫洞

克尔-纽曼度规

克尔-纽曼黑洞

克尔-纽曼解

克莱茨曼标量

克莱因-戈登场

L

雷查德利方程

雷斯勒-诺斯特朗姆度规

雷斯勒-诺斯特朗姆黑洞

雷斯勒-诺斯特朗姆解

类光测地线

类光一般性条件

类空测地线

类空超曲面

类时测地线

类时一般性条件

黎曼-彭罗斯猜想

黎曼联络

李导数

理想尘埃

理想流体

量纲分析法

量子不等式

量子涨落

临界参数

零尘埃

零能量条件

罗伯逊-沃尔克度规

裸奇点

洛伦兹度规

洛伦兹因子

M

闵科夫斯基度规

N

能量动量张量

能量条件

P

彭罗斯不等式

彭罗斯猜想

膨胀标量

平均能量条件

平均外曲率

普朗克常数

Q

奇异物质

强渐近可预测

强能量条件

强因果性条件

切变张量

曲率张量

全局方法

全局双曲

R

弱能量条件

弱宇宙监督假设

S

萨根式问题

生命传输机

施瓦西度规

施瓦西黑洞

施瓦西解

施瓦西奇点

时间机器

时序保护假设

时序条件

事件视界

随动观测者

T

拓扑改变

W

外曲率

外森比克公式

未来柯西视界

未来柯西展开

涡旋张量

无限发达的文明

Y

雅可比场

一般性条件

因果性条件

引力微子

诱导度规

宇宙监督假设

宇宙学奇点

Z

张力

正则形式

正质量定理

终极理论

逐点能量条件

主能量条件

祖父悖论

最大柯西展开

人名索引

A

阿诺维特 (Richard Arnowitt)
阿什提卡 (Abhay Ashtekar)
艾里斯 (George Ellis)
爱丁顿 (Arthur Eddington)
爱因斯坦 (Albert Einstein)
奥本海默 (Robert Oppenheimer)

B

贝林斯基 (Vladimir Belinski)
玻德 (Dirk Polder)
玻尔 (Niels Bohr)
布雷 (Hubert Bray)
布里尔 (Dieter Brill)

D

戴舍 (Stanley Deser)
德维外迪 (Indresh Dwivedi)

F

弗里德曼 (Alexander Friedmann)
福特 (Larry Ford)

G

格里沙鲁 (Marc Grisaru)

H

哈尔彭 (Paul Halpern)
哈梅迪 (Rufus Hamadé)
哈特尔 (James Hartle)
赫尔斯 (Russell Hulse)
怀廷 (Bernard Whiting)
惠勒 (John Wheeler)
惠斯肯 (Gerhard Huisken)
霍金 (Stephen Hawking)
霍罗威茨 (Gary Horowitz)

J

吉本斯 (Gary Gibbons)
杰罗奇 (Robert Geroch)

K

卡拉特尼科夫 (Isaak Markovich Khalatnikov)
卡斯丹 (Jerry Kazdan)
卡西米尔 (Hendrik Casimir)
凯 (Bernard Kay)
克拉克 (Arthur Clarke)
克里斯多洛 (Demetrios Christodoulou)

L

拉曼 (Chandrasekhara Raman)
拉莫理奥克斯 (Steve Lamoreaux)
莱特 (Maria Leite)
朗道 (Lev Landau)

勒梅特 (Georges Lemaître)
雷波维茨 (Clement Leibovitz)
雷查德利 (Amal Raychaudhuri)
栗弗席兹 (Evgeny Lifshitz)
罗伯逊 (Howard Robertson)
罗曼 (Thomas Roman)
罗森 (Nathan Rosen)
洛卡姆 (Joachim Lohkamp)

M

马斯登 (Jerrold Marsden)
米斯纳 (Charles Misner)
莫勒 (Christian Møller)
莫里斯 (Mike Morris)
莫里斯奇 (Osvaldo Moreschi)

N

纳里卡 (Jayant Narlikar)

P

帕克 (Thomas Parker)
佩奇 (Don Page)
彭罗斯 (Roger Penrose)
普赖斯 (Richard Price)
普雷斯基尔 (John Preskill)

Q

钱德拉塞卡 (Subrahmanyan Chandrasekhar)

乔希 (Pankaj Joshi)
丘成桐 (Shing-Tung Yau)

S

萨根 (Carl Sagan)
塞弗特 (Hans Seifert)
森 (Ashoke Sen)
舍恩 (Richard Schoen)
施奈德 (Hartland Snyder)
施瓦西 (Karl Schwarzschild)
斯帕林 (George Sparling)
斯图尔特 (John Stewart)
索恩 (Kip Thorne)
索金 (Rafael Sorkin)

T

泰勒 (Joseph Taylor)
泰特伯姆 (Claudio Teitelboim)
陶布斯 (Clifford Taubes)
梯普勒 (Frank Tipler)
托曼 (Richard Tolman)

W

威顿 (Edward Witten)
威尔斯 (H. G. Wells)
维希外希瓦拉 (C. V. Vishveshwara)
沃尔德 (Robert Wald)
沃尔克 (Arthur Walker)

伍尔加 (Eric Woolgar)

X

肖盖 (Yvonne Choquet-Bruhat)

肖普推克 (Matthew Choptuik)

Y

伊尔玛能 (Tom Ilmanen)

伊斯雷尔 (Werner Israel)

尤瑟福 (Ulvi Yurtsever)

约德杰斯 (Peter Yodzis)

约尔当 (Pascual Jordan)

参考文献

综合读物

1. Ciufolini I, Wheeler J A, Gravitation and Inertia [M]. Princeton: Princeton University Press, 1995.
2. Hawking S W, Ellis G F R. The Large Scale Structure of Space-Time [M]. Cambridge: Cambridge University Press, 1975.
3. Joshi P S. Global Aspects in Gravitation and Cosmology [M]. Oxford: Oxford University Press Inc., 1993.
4. Misner C W, Thorne K S, Wheeler J A. Gravitation [M]. New York: W. H. Freeman and Company, 1973.
5. Wald R M. General Relativity [M]. Chicago: University of Chicago Press, 1984.
6. Witten L. Gravitation: An Introduction to Current Research [M]. New York: Wiley 1962.
7. 刘辽, 赵峥. 广义相对论[M]. 北京: 高等教育出版社, 2004.

能量条件

1. Barcelo C, Visser M. Twilight for the energy conditions? [J]. Int. J. Mod. Phys, 2002, D11: 1553-1560.
2. Ford L H. The Classical Singularity Theorems and their Quantum Loopholes [J]. Int. J. Theor. Phys., 2003, 42: 1219-1227.

奇点与奇点定理

1. Foertsch T, Wegner B. Selected Topics in Singularities of Space-time, Proc. Summer School on Diff [C]. Geometry-Dep. de Matematica, Universidade de Coimbra-September 1999: 39-52.
2. Hawking S W, Penrose R. The Nature of Space and Time [M]. Princeton:

Princeton University Press, 1996.

3. Natário J. Relativity and Singularities-A Short Introduction for Mathematicians [J]. arXiv: math/0603190[math.DG].

4. Schoen R, Yau S T. The Existence of a Black Hole Due to Condensation of Matter [J]. Comm. Math. Phys., 1983, 90: 575-579.

5. Senovilla J M. Singularity Theorems in General Relativity: Achievements and Open Questions [J]. arXiv: physics/0605007[physics.hist-ph].

正质量定理

1. Bray H L. Proof of the Riemannian Penrose Inequality Using the Positive Mass Theorem [J]. J. Diff. Geom., 2001, 59: 177-267.

2. Bray H L, Lee D A. On the Riemannian Penrose Inequality in Dimensions Less Than 8 [J]. arXiv: 0705.1128 [math.DG] .

3. Frauendiener J. Conformal Infinity [J]. Living Rev. Relativity, 2004, 7: 1.

4. Gibbons G W, et al. Positive Mass Theorems for Black Holes [J]. Comm. Math. Phys., 1983, 88: 295-308.

5. Kazdan J L. Positive Energy in General Relativity [J]. Séminaire N. Bourbaki, 1982-1983, 25: 315-330.

6. Lohkamp J. The Higher Dimensional Positive Mass Theorem I [J]. math.DG/0608795.

7. Moreschi M, Sparling G A J. On the Positive Energy Theorem Involving Mass and Electromagnetic Charges [J]. Comm. Math. Phys., 1984, 95: 113-120.

8. Parker T, Taubes C H. On Witten's Proof of the Positive Energy Theorem [J]. Comm. Math. Phys., 1982, 84: 223-238.

9. Penrose R, Sorkin R D, Woolgar E. A Positive Mass Theorem Based on the Focusing and Retardation of Null Geodesics [J]. gr-qc/9301015.

10. Schoen R, Yau S T. On the Proof of the Positive Mass Conjecture in

General Relativity [J]. Comm. Math. Phys., 1979, 65: 45-76.

11. Witten E. A New Proof of the Positive Energy Theorem [J]. Comm. Math. Phys., 1981, 80: 381-402.

12. 丘成桐, 孙理察. 微分几何讲义[M]. 北京: 高等教育出版社, 2004.

宇宙监督假设

1. Choptuik M W. Universality and Scaling in Gravitational Collapse of a Massless Scalar Field [J]. Phys. Rev. Lett., 1993, 70: 9.

2. Harada T, Iguchi H, Nakao K. Physical Processes in Naked Singularity Formation [J]. Prog. Theor. Phys., 2002: 107: 449.

3. Joshi P S. Cosmic Censorship: A Current Perspective [J]. Mod. Phys. Letts., 2002, A17: 1067.

4. Mars M. An Overview on the Penrose Inequality [J]. Journal of Physics: Conference Series, 2007, 20: 012004.

5. Penrose R. The Question of Cosmic Censorship [J]. J. Astrophys. Astr., 1999, 20: 233-248.

6. Schoen R, Yau S T. The Existence of a Black Hole due to Condensation of Matter [J]. Comm. Math. Phys., 1983, 90: 575-579.

7. Singh T P. Gravitational Collapse, Black Holes and Naked Singularities [J]. arXiv: gr-qc/98050661.

8. Wald R M. Gedanken Experiments to Destroy a Black Hole [J]. Ann. Phys., 1974, 83: 548-556.

9. Wald R M. Gravitational Collapse and Cosmic Censorship [J]. arXiv: gr-qc/9710068.

虫洞物理学

1. Bordag M (eds). The Casimir Effect 50 Years later [M]. Singapore: World Scientific, 1999.

2. Dalvit D, Milonni P, Roberts D, Rosa F (eds). Casimir Physics [M]. Berlin Heidelberg: Springer, 2011.
3. Ford L H, Roman T A. Quantum Field Theory Constrains Traversable Wormhole Geometries [J]. Phys. Rev., 1996, D53: 5496-5507.
4. Halpern P. Cosmic Wormholes: The Search for Interstellar Shortcuts [M]. New York: Plume, 1993.
5. Hawking S W. Hawking on the Big Bang and Black Holes [M]. Singapore: World Scientific, 1993.
6. Lamoreaux S K. Demonstration of the Casimir Force in the 0.6 to 6 μm Range [J]. Phys. Rev. Lett. 1997, 78: 5.
7. Milton K A. The Casimir Effect: Physical Manifestations of Zero-Point Energy [M]. Singapore: World Scientific, 2001.
8. Morris M S, Thorne K S. Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity [J]. Am. J. Phys. 1988: 56: 395.
9. Roman T A. Some Thoughts on Energy Conditions and Wormholes [J]. arXiv: gr-qc/0409090.
10. Thorne K S. Black Holes & Time Warps [M]. New York: W. W. Norton & Company, 1995.
11. Visser M. Lorentzian Wormholes: From Einstein to Hawking [M]. Woodbury: AIP Press, 1996.

后 记

终于到了可以写后记的时候，这大概是写一本书的过程中最轻松的时刻吧。这本书前后拖了七年多，跟上一本书——《黎曼猜想漫谈》——相比虽还算“高效”，也终究够得上写点回忆了——毕竟，人生能有几个“七年多”呢？

本书的缘起可以追溯到写科普的起点：2002年。那一年，我开始撰写一个题为星际旅行漫谈的科普系列。那个系列承袭了我在天文和科幻两方面的兴趣，是我最早的科普。其中有关虫洞的部分则可以说是本书的先声。记得很早的时候——也许是还在念中学的那会儿吧——我曾从一本科普书上读到过一位科学家（或科普作家）的话，大意是：光速是有限的，我们的子孙后代将在极度的孤寂中体会它的深意。那句话给我的印象非常深（可惜却记不起出处了，只能把惭愧推给年岁），在撰写星际旅行漫谈时，也常常会想起它来。我喜爱物理，但物理规律有时连我自己也觉得有些冷——或者所有真正的规律都是冷峻的吧，想到浩瀚星空中被光速极限无情分隔开的遥远部分，是我们几乎永远也不可能有机会前往的，不免有些惆怅。

在那样的背景下，虫洞这种似乎能变相突破光速极限的手段引起了我的极大兴趣。而且跟其他一些科幻设想——比如阿西莫夫笔下的“跃迁”（jump），或某些天知道怎么造出来的“超光速飞船”——相比，虫洞的物理色彩要浓厚得多，甚至可以算是有一定理论根基的，这对于我这种既喜欢科幻，又是物理专业的人来说，是格外有吸引力的。当然，如果您已经读过本书，那么也许已经跟我一样不无遗憾地注意到了，依据我们现在所知的物理规律，虫洞作为星际旅行的通道，实现的可能性是微乎其微的。作为科幻爱好者，我期盼正面消息，甚至梦想着在自己有生之年就能听到那样的消息；但作为科普作者，我只能忠实于现代物

理，那是确立科普作品价值的核心坐标系，也是它有别于幻想的存在理由。当然，凡存在显著争议的部分，我都尽量给出了提示——不仅为了严谨，也是因为那是某种程度的希望所系，是我自己所乐意见到的。

在虫洞这个题材上，我虽然已写过文章，但前面提到的星际旅行漫谈中有关虫洞的部分，以及后来发表在《科幻世界》上的“虫洞：旅行家的天堂还是探险者的地狱？”（《科幻世界》2006年第5期，科幻世界杂志社）都是普通科普。而要想把虫洞介绍得透彻些，显示出它在现代物理中的理论根基，则必须采用所谓专业科普的形式，这使我萌生了撰写本书的念头——也就是本书的缘起。

不过，撰写本书的念头虽来自对虫洞的兴趣，本书的题材却并不仅限于虫洞。因为在收集素材的过程中，我注意到在虫洞物理学中有一组条件起着很重要的作用，那就是能量条件，而能量条件在其他一些广义相对论课题——比如奇点与奇点定理、正质量定理、宇宙监督假设——中也有重要应用，那些题材也很有意思，并且大都是普通广义相对论教材中不介绍或介绍得不深入，从而对专业科普来说可算是好课题的。于是我决定扩展写作范围，以能量条件为线索，对那些课题一并进行介绍。由于这一原因，本书最初在我的网站上连载时，所用的标题乃是“现代物理中的能量条件”。同时，也是由于这一原因，本书的篇幅才膨胀为了“书”。

本书的撰写始于2005年12月，前两章的内容（即“能量条件”和“奇点与奇点定理”）曾于2007—2008年间在《现代物理知识》上连载。可惜，该连载只进行了两章，第3章（即“正质量定理”）因“不够通俗”而遭审稿编辑“枪毙”，整个连载遂告“腰斩”。当时我网站上的连载已进行到了第4章（即“宇宙监督假设”），由于约稿压力的终止，在2008年9月写完该章之后，我就停了下来，将写作兴趣转向了其他方面。于是本书也像《黎曼猜想漫谈》一样，一度成为了“烂尾楼”。

本书的重新启动是在2012年4月，当时《黎曼猜想漫谈》已经写

完，在与清华大学出版社的编辑讨论下一本书的计划时，我提到了这个已经完成了大半的题材。编辑表示感兴趣，并寄来了出版合同。于是自2012年12月起，我开始撰写本书的最后一章（即“虫洞物理学”），并于2013年3月完成了全书。

本书作为我的第四本书，已不再有出版第一本书那样的新鲜感，但它的完成仍有一个比较特殊的意义，那就是填补了一个不应有的空白。有些读者也许知道，我这位写过两本天文书、一本数学书的“非著名科普作家”其实是物理专业的。一个物理专业的人，写了三本书却没有一本是关于物理的，多少有些愧对专业，也有些缺憾。本书的完成可以算是弥补了缺憾，从这点上讲，它的出版对我有一种不同于前三本书的欣慰感。

当然，真正的欣慰不是我写了一本物理书，而应该是这本物理书能得到部分读者的喜爱——这一点我知道是不容易的，起码要比前三本书都更难，因为与同为专业科普的《黎曼猜想漫谈》不同，本书没有多少科学家的逸事，也没有什么戏剧性的情节，这对于博取读者的喜爱是非常不利的。但我希望，本书凭借“事后诸葛”特有的时间优势，能做到在叙述逻辑上比原始论文更顺畅或更易被理解，以及对后续发展有更清晰的视野。若如此，则它对那几个广义相对论专题可以起到导读作用。我并且希望，起码有一部分读者能从这样的导读中受益。

在结束这篇后记之前，有一点关于笔误的小小感想要在这里提一下。众所周知，在所有错误中，笔误是相对轻微的，往往读者自己就能分辨。但专业科普由于带有数学公式，笔误一旦出现在公式中，影响就会变大，且看上去不像笔误倒像硬伤。日本物理学家汤川秀树写过这样一句体会：“拿到出版的书后，每次都会有很多问题闯入我的眼帘，我会为此而坐卧不安。”我曾经觉得自己的脸皮要比汤川先生厚，不至于坐卧不安的，但后来在“豆瓣读书”看到一位读者针对《黎曼猜想漫谈》中一处公式笔误的留言“才看几页，就发现了公式错误，原本很期待的

一本书，很怀疑严谨性了”时感到脸上发热，才不无遗憾地意识到自己的脸皮其实也不厚。而更遗憾的是，那公式在我的网站上其实是正确的，纯属整理文稿时的疏漏——那种感觉，宛如昔日大考之后发现一个粗心大意的错误。为了减少笔误的影响，我在这里公布一下自己的网址，以便读者发现笔误或疑似笔误时，可以核对网络版，也可以发邮件告诉我。任何笔误一经确认，都将第一时间在网络版上更正，并将等待重印或再版的机会在书面版上更正。

最后，谢谢诸位的阅读，并请大家相信一位有着几十年读书史和十几年爬网史的“书虫”兼“网虫”的经验：阅读实体书的感觉绝对要比阅读网络版好得多。

作者网址：<http://www.changhai.org/>

2013年3月30日写于纽约



理解科学丛书·卢昌海科普著作



原点阅读

清华大学出版社

原点阅读（The Origin），清华大学出版社旗下的图书品牌，秉承“科学，让个人更智慧，让社会更理性”的理念，致力于科学普及和科技文化类图书的出版，传播科学知识、科学精神、科学方法，展现科学的真实、独立、智慧、多变、宽容、动人及迷人。

理解科学丛书·卢昌海科普著作

10 BILLION YEARS
FROM PAST TO
FUTURE
The Story
of the Sun
上下百亿年
太阳的故事
卢昌海◎著

重返古希腊 天文自助游
阳光里的奥秘 大气里的谜团
从地心说到日心说 从日全食到相对论
光子的逃亡 太阳的脉搏
小小的粒子 大大的模型
太阳的过去、现在和未来

清华大学出版社



理解科学丛书

10 BILLION YEARS
FROM PAST TO
FUTURE
The Story
of the Sun

上下百亿年 太阳的故事

卢昌海◎著

清华大学出版社
北京

目 录

[内容简介](#)

[序言](#)

[1 重返古希腊](#)

[2 天文自助游：推算太阳的大小和远近](#)

[3 地心说vs日心说](#)

[4 日食——既寻常又稀有的奇观](#)

[5 插曲：爱丁顿在1919](#)

[6 阳光里的奥秘](#)

[7 物理自助游：推算太阳的质量、光度和表面温度](#)

[8 光明的源泉 恐怖的核心](#)

[9 细小的粒子 巨大的谜团](#)

[10 标准太阳模型vs粒子物理标准模型](#)

[11 光子大逃亡](#)

[12 太阳的脉搏](#)

[13 谜团锦簇的太阳大气层](#)

[14 太阳的过去和未来](#)

[附录 太阳档案](#)

[人名索引](#)

[术语索引](#)

[参考文献](#)

[返回总目录](#)

图书在版编目 (CIP) 数据

上下百亿年：太阳的故事/卢昌海著.--北京：清华大学出版社，2015

（理解科学丛书）

ISBN 978-7-302-40762-1

I. ①上... II. ①卢... III. ①太阳—青少年读物 IV. ①P182-49

中国版本图书馆CIP数据核字（2015）第161102号

责任编辑：邹开颜

封面设计：

责任校对：王淑云

责任印制：

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社总机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印刷者：

装订者：

经 销：全国新华书店

开 本：165mm×240mm 印张：12.5 彩页：2

字 数：179千字

版 次：2015年8月第1版

印 次：2015年8月第1次印刷

印 数：1～ 000

产品编号：

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

内容简介

本书将带读者从一次虚拟的古希腊之旅开始，逐步深入地了解太阳的奥秘，以及人类为探索这些奥秘所付出的艰辛和努力，所获得的成果和教训。除介绍知识外，本书还穿插了少许章节，引导读者运用逻辑推理及初等几何知识，再现古希腊先贤们的某些精彩推理，以及运用含义简明的物理学原理，推算太阳的某些物理参数，亲身体验科学探索的感觉。

本书融语言的生动风趣与内容的严谨翔实于一体，让读者在享受阅读愉悦的同时学到丰富的知识，并体验科学探索的严谨及科学发现的激动人心。本书适合广大天文爱好者及大、中学生阅读。

序言

《太阳的故事》初版于2011年，是我的第二本书，也是我迄今所有书中最受冷遇的一本。如今，清华大学出版社以不怕亏损的绝大勇气为此书出修订版，本该“避亲”的我也不便袖手旁观，就厚颜为此书“吆喝”几句吧。

《太阳的故事》虽然书名超土，题材也不稀有，但在我眼里，却是比其姊妹篇《寻找太阳系的疆界》[\[1\]](#)写得更用心力并且也更好的。为什么呢？因为后者是单纯的“历史题材”，而且是“寻找太阳系的疆界”这单一领域内的历史，受到惯常的全局性时间顺序的限制。昔日鲁迅的《中国小说史略》被诬为抄袭时，鲁迅在回应时说过这样一句话：“自然，大致是不能不同的，例如他说汉后有唐，唐后有宋，我也这样说，因为都以中国史实为‘蓝本’。我无法‘捏造得新奇’。”像《寻找太阳系的疆界》那样的题材也有类似的尴尬：别人依天王星、海王星、冥王星的顺序写，我也只能这样写。不仅如此，由于“寻找太阳系的疆界”是比“中国小说史”还单一得多的领域，就连每个行星的发现者、发现过程等也都没得挑拣，更遑论“捏造得新奇”。因此对作者来说，虽可作——并且也作了——文笔上的发挥、细节上的考辨，以及局部结构上的规划等，总体的写作自由度是比较小的，相应地，独特性也就比较低。

《太阳的故事》则不同，它侧重于历史但不是单纯的“历史题材”，而是科普与科学史的融合；而且它的历史部分也并非单一领域内的历史，而是涉及天文和物理的多个分支。这种复杂性给了作者很大的写作自由度。作为结果，《太阳的故事》的整个结构——以“穿越”而始，由

几何至物理，从核心到外围，最后以太阳“通史”为终——都是我自己规划，并且自认为比较独特的。这其中以《重返古希腊》为开篇的“穿越”式写法更是我很久以前就萌生的一个写作思路的贯彻，在我自己是颇为珍视的。另外，从内容上讲，《太阳的故事》涉及的知识点比《寻找太阳系的疆界》多得多，对科学方法和科学探索过程的介绍也深入得多；在写作过程中，则参阅和甄选了多得多的资料。这些对作者来说都是更用心力并且也更好的地方。

因此，《太阳的故事》出修订版是我非常乐意见到的。

不过另一方面，我也必须坦白，这修订版所“修订”的基本上也就只是“版”，以内容而论实在是微不足道的（因此已有初版的读者不必再次破费）。这一点其实是我所有作品的修订版共有的特点。之所以如此，主要是——如我在《小楼与大师：科学殿堂的人和事》^[2]——书的自序中所述——由于我写作速度较慢，“使得写作过程往往长到了对题材的兴趣将尽而书稿远未完成的程度”。由此造成的后果，则是书稿的完成之日，往往也就是兴趣的透支极限，从而在很长的时间之内都不会再对同一题材感兴趣，更不会去修订。从某种意义上讲，一本书出版后，在我眼里它本身也就成为了历史，除订正笔误外，在文字上我一般是让它维持原貌的。对这种有懒惰之嫌的做法，容我略作辩白：我作为一个喜欢买书的人，屡次吃过大幅变更内容的“修订版”的亏。比如有一年我买了一套有关民国大师的三卷本的“大书”，还没高兴太久，就听说该书出了修订版，“新增10万字全新内容”，让“故事真正完整”。那一刻的心情，固然有一分是叹服作者的勤奋，却有九分是为自己的版本缺了“10万字全新内容”，及故事的并非“真正完整”而沮丧。我希望，喜欢我作品的读者无需在短短几年间就尝到同样的沮丧——当然，前提是作品本身不会在那样的时间内过时，而这，我寄望于写作时付出过的心力，以及科

学史这个我所侧重的领域本身的相对稳定。

最后，让我以对修订之处的罗列，来结束自己的“吆喝”吧：

（1）增添了一篇“序言”（即本文）。

（2）订正了几处笔误（大都为措辞、标点、译名的微调）。

（3）变更了书名——《太阳的故事》这个超土的书名作为本书受冷遇的替罪羊，遭到了编辑的撤换。我效仿昔日的历史畅销书《上下五千年》，提议了《上下百亿年》这一新书名。

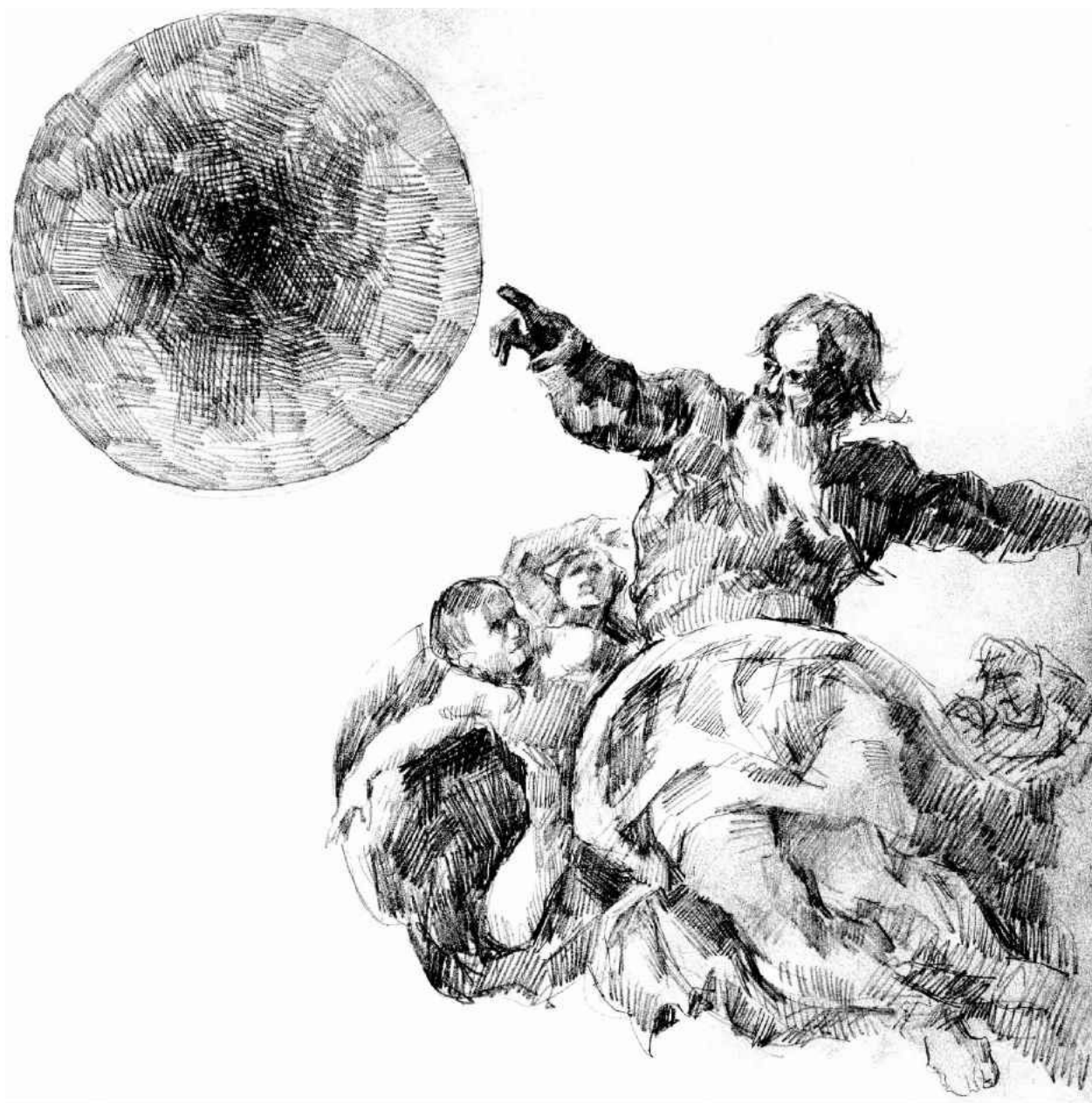
（4）增添了一些手绘插图。

（5）增添了人名和术语索引。

以上就是对这本修订版的说明，但愿有更多的读者注意到并喜欢上本书。

[\[1\]](#)《寻找太阳系的疆界》由清华大学出版社初版于2009年，修订版更名为《那颗星星不在星图上：寻找太阳系的疆界》，于2013年出版。

[\[2\]](#)《小楼与大师：科学殿堂的人和事》由清华大学出版社出版于2014年。



绘画：张京

1 重返古希腊

说到天文学，很多人的眼前都会浮现出深邃的天幕和宝石般闪亮的星辰。其实，在我们这个小小星球上所能看到的最显眼的天文现象并不在黑夜，而是在白天。

在每一个晴朗的白天，天空中都挂着一个极为显眼的天体：太阳。

对于像太阳这样显眼的天体，我们当然不必问它是什么时候被发现，以及怎样被发现的，因为那显然跟人类本身同样古老，跟睁开眼睛同样直接。但是，除了这两个不必问的问题外，有关太阳的其他问题可就大都不是省油的灯了，有些甚至直到今天也没有确切答案。不过虽然不“省油”，它们点亮的却是人类的智慧。从某种意义上讲，对这个天空中最显眼的天体的持续探索，对那些“不省油”的问题的认真求解，是人类从睁开眼睛看世界，到逐渐理解世界的某些方面所走过的几千年漫长路的一个缩影。

现在就让我们从那些问题当中最简单的两个说起吧：太阳有多大？它离我们有多远？

这两个问题的答案，在今天也许已是很多小学生都知道的常识——不就是两个数字嘛。但是，这两个问题的答案果真只是两个数字吗？让我们来作这样一个设想，假设我们用时间机器把一位知道这两个数字的小学生送回人类文明发源地之一的古希腊。我们想知道的是：这位来自21世纪的小学生能做什么？

显然，单以某些知识——比如有关太阳有多大和离我们有多远的这

两个数字——而论，他（她）已经远远胜过了古希腊的任何一位先贤。但我们会在从古希腊流传下来的史书中读到有关这位博学强记的小学生的故事吗？他（她）能凭那些博学强记的知识就成为令后世之人高山仰止的先贤中的一位吗？我想答案是否定的。原因很简单，知道两个数字和让别人理解并信服那两个数字是完全不同的事情，后者恐怕不是每一位小学生能够做到的。如果仅仅能说出两个数字，却无法让别人理解和信服，那只会被当成信口开河，而不会被载入史书。

那么，假如不是小学生，而是本书的读者，您有幸（或不幸）被送回到了古希腊，您有办法让那些喜爱思考的古希腊先贤们相信那两个数字，相信天空中那个看上去只有贝壳大小的太阳其实是肚子里能装下一百多万个地球的庞然大物吗？或者换一个说法：若是您被送回到了古希腊，却忘记了那两个数字，您有办法凭自己的能力，以一种令人信服的方式重新找回它们吗？或者更一般地，如果您站在了古希腊的天空下，却忘记了所有的天文知识，您能凭借自己的能力找回其中的多少呢？

让我们就从这个假想的问题开始重温一下人类智慧的启蒙时代，并从那里开始讲述我们有关太阳的故事吧。

要想找回已被忘记的天文知识，您要做的第一件事情显然就是仰望天空，因为那里——并且只有那里——才是天文知识的直接来源。如果您的仰望天空只是偶一为之，您也许会觉得天上的日月星辰都是静止的，因为它们当中没有一个会像飞鸟一样在一瞥之下就让人察觉它们的移动。但即便如此，您也会在一天之内就发现太阳的东升西落，因为它直接影响到周围环境的明暗和冷暖。要发现月亮的运动也很容易，因为在任何一个有月亮的夜晚，您仰望天空时都很难不注意到这个独一无二的天体，而一旦注意到它的存在，那么在下次仰望天空时，就很难不注意到它的位置变化。

对一般人来说，自己所能发现的天文知识也许就到此为止了。天上除日月之外虽然还有很多星星，星星虽然也和日月一样东升西落，但一个视力良好的人在一个晴朗的夜晚所能看到的星星有几千颗之多，若非特别留意，除了有一种繁星似尘的感觉外，恐怕是不会对其中任何一颗星星留下具体印象的。而如果没有对任何一颗星星留下具体印象，那么在下次仰望天空时就很难注意到它们的移动。

要想找回尽可能多的天文知识，您当然不能像一般人那样过目就忘。为了研究星星的运动，您开始进行细致的观测，并对不同时刻每颗星星的位置进行记录。您很清楚，观测越细致，记录越详尽，有可能找回的天文知识就越丰富。由于在苍穹之上缺乏参照，不易度量位置或角度，您也许会想到在地上立一些固定的物件作为参照，如果手下有一些可以使唤的人的话，您也许还会想要设计建造一些更复杂的参照物，那些东西若是建得足够牢固，以至于能一直保留到今天的话，就会变成重要的历史遗迹：古观象台。

无论您的记录详尽还是粗略，只要记录了，哪怕只记录几天，您也会发现所有的星星都和日月一样东升西落。用后世的术语来说，这是天体的周日视运动（**apparent diurnal motion**），如图1.1所示。由此您也许还会进一步总结出一个规律，那就是日月星辰都在围绕着地球转动。在历史上，这是著名的地心说（**geocentric model**），它后来受到宗教势力的维护，成为垄断天文界长达两千年的正统理论。随着观测数据的积累，以后您会发现很多理由让您放弃这一理论。它后来也的确被放弃了。在某些后世之人的眼里它甚至有些声名狼藉（那其实是宗教惹的祸）。但在一开始，在只有粗略观测数据的年代里，它是一种既符合观测数据，又符合直觉的理论。您有理由为发现这一理论而自豪。周日视运动的发现也意味着您已经发现了“日”这个时间计量单位，它是周日视

运动的周期，也可以说是昼夜交替这一粗糙周期概念的精细版。[\[1\]](#)



图1.1 星星的周日视运动

当您的天文观测坚持到几十天时，除了周日视运动外，您还会注意到另一种很重要的天文周期现象，那就是月相（**phases of the moon**）的变化（彩图2）。与太阳总是圆的，以及星星总是像一个点不同，月亮这个夜空中最显眼的天体在不同日子里会呈现不同的形状：有时是满月，有时是半月，有时则是弯月。这种变化被称为月相的变化，它大约每隔29.53天重复一次。注意到这种有趣而美丽的周期现象，意味着您发现了“朔望月”（**synodic month**）这一时间计量单位。很多早期的文明都曾用过这一时间计量单位，直到今天它仍有一定的应用，是阴历（**lunar calendar**）这一历法的基础。[\[2\]](#)

当您的天文观测坚持到十几个月时，除了周日视运动和月相的变化

外，您还会发现一种更缓慢的天文周期现象。您会注意到在太阳升起和落下的时候，天空中依稀可见的那些星星的位置在一天天缓慢地改变着。这种缓慢改变的逐渐积累，使得在不同的季节里，伴随太阳升起和落下的星星是不同的。这说明什么呢？说明太阳在背景星空中的位置不是固定的，除了周日视运动夕卜，它还参与了一种更加缓慢的运动。仔细的观测表明，那种运动大约每隔365.24天重复一次，它既沿东西方向，也沿南北方向，与周日视运动所在的平面有一个 23.4° 左右的夹角，这个夹角决定了太阳在冬天和夏天所能到达的最大纬度——即南北回归线的纬度。注意到了那种运动，意味着您发现了所谓的太阳周年视运动（**apparent yearly motion**）以及“年”这一时间度量单位，后者是太阳周年视运动的周期，也可以说是四季变化这一粗略周期的精细版。^[3]

您不知疲倦地坚持着自己的天文观测，当头发都快花白了的时候，您在天空中又发现了一些更微妙的运动。您会发现在那看起来彼此相似的满天繁星之中，有五颗星星的位置与日月一样相对于背景星空在缓慢地移动着，其中有几颗星星的移动方式还相当复杂，比如有时会停止，有时还会逆行。如果您发现了这些被后人称为行星表观视运动（**apparent motion of planets**）的现象，那表明您已经发现了金、木、水、火、土五大经典行星。除了这些发现以外，在经年累月的观测中您还会偶尔发现一些流星和彗星，并观测到一些日食和月食。

在古希腊的条件下，您自己所能从事的天文观测大致就是这些。不过，假如您能有幸找到一些前人留下的观测记录的话，您也许能通过将彼此的记录相互比较，而发现一种在自己的有生之年里很难单独发现的东西，即周日视运动的轴线本身的缓慢转动。这种转动的周期约为25 800年。这一现象用后世的术语来说就是所谓的地球自转轴的进动（**precession of the Earth's rotation axis**）。在它的影响下，因距离北天极

（即周日视运动的轴线北端）很近而被称为北极星（Polaris）的小熊座 α 星（ α Ursa Minor）在几千年后将会失去北极星这一光荣称号。

完成了上面这些观测发现，您就不仅凭借自己的能力赶上了古希腊先贤们在观测天文学上曾经达到过的水准，而且也基本上穷尽了17世纪之前天文学上几乎所有重要的观测发现。罗列起来似乎不难，做起来却不无艰辛。在不知不觉间，您这位来自21世纪的人，已几十年如一日地将古希腊人的天文事业当成了自己的事业。（这是一种什么精神？）

不过，这些天文发现虽然了不起，却还不足以让您被写入史书。因为眼睛是人人都有，很多勤奋的普通人——其中既有古希腊人，也有其他古文明国度的人——也能作出同样的发现。真正将智者区别于普通人的除了勤奋，还有智慧。例如几何与推理的能力，这种能力无疑是一种智慧。现在您就要用自己的智慧来做一些单纯的天文观测无法做到的事情。比方说，您要寻找前面提到过的那两个数字：太阳的大小以及它离我们的远近。

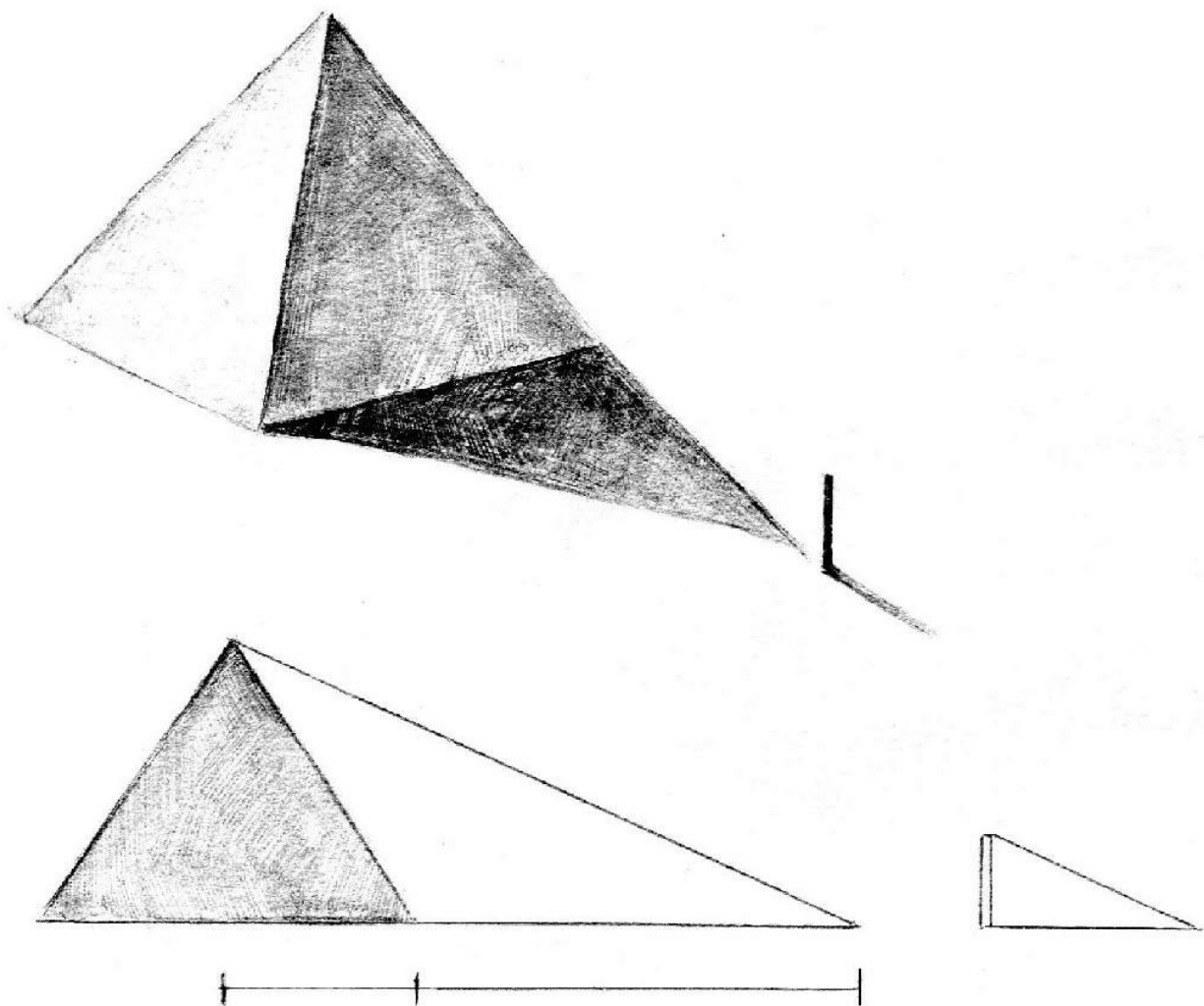
没有谷歌（Google），没有百度（Baidu），而且也没“病”（Bing），您有办法自己找出那两个数字吗？

[\[1\]](#)从周日视运动的周期中衍生出的“日”的概念其实不止一种：由同一颗星星（行星除外）在两个相邻夜晚经过天空中同一个位置的时间间隔所定义的“日”称为恒星日（sidereal day）；由太阳在两个相邻白天经过天空中同一个位置的平均时间间隔所定义的“日”则称为平均太阳日（mean solar day）。由于后文即将提到的太阳周年视运动的影响，平均太阳日比恒星日长了约3分56秒（感兴趣的读者可以用本节给出的数据自行推算一下这两种“日”的差异）。由于太阳与我们日常

生活的关系远比星星密切，我们在普通日历中所用的“日”是指平均太阳日。细心的读者可能会问：平均太阳日中的“平均”二字是什么意思？那是指将地球公转轨道等效为一个平均圆轨道，以避免“日”的长短受地球公转轨道的椭圆性影响。当然，不作那样的平均也可以谈论“日”这个概念，那样的“日”被称为表观太阳日（apparent solar day），它的长短会随季节而变。

[2]朔望月这一中文名称中的“朔”指的是新月，“望”指的是满月。要注意的是，朔望月只是月相变化的周期，而不是月球绕地球公转的周期，后者是所谓的恒星月（sidereal month），只有27.3天（感兴趣的读者可以用本节给出的数据自行推算一下这两种“月”的差异）。另外，朔望月不同于以太阳周年视运动为基础的阳历（solar calendar）中的月。之所以不同，是因为朔望月并不恰好等于阳历中一年的十二分之一，如果我们用它来表示“月”，就无法与“年”合拍，由此会导致很多不方便之处，比方说北半球的7月就无法稳定地对应于夏天（因为当“年”和“月”的不合拍累积到六个月时，它就会变成冬天）。不过有得就有失，阳历中的“月”虽然保证了与“年”的合拍，却失去了表示月相的作用，比方说“中秋月圆”在阳历中就没有一个固定的日子。由于太阳与我们日常生活的关系远比月亮密切，我们日常所用的“月”是指阳历中的月。

[3]确切地说，这个“年”是所谓的回归年（tropical year），它比地球绕太阳的公转周期，即所谓的恒星年（sidereal year）短了约20.4分钟，这两者的差异是由后文即将提到的地球自转轴的进动造成的（感兴趣的读者可以用本节给出的数据自行推算一下这两种“年”的差异）。我们在阳历中所用的“年”是指回归年。



绘画：张京

2 天文自助游：推算太阳的大小和远近

虽然您要寻找的数字有两个，但很清楚，实际上只要找到其中一个就行了。因为太阳就在天上，它看起来有多大您早就知道了，它的真实尺寸越大，意味着离我们越远，反之，真实尺寸越小，意味着离我们就越近。这表明，在太阳的大小和远近这两者之间存在完全确定的关系，只要知道任何一者，就可以推算出另外一者。

那么，在大小和远近这两者之间您该选择哪一者入手呢？从兴趣上讲，您也许会对大小更感兴趣，因为那才是属于太阳本身的性质，但实际上，您却只能从远近入手。对于普通物体来说，这两种选择并无多大分别，只要用一把尺子，您爱测量哪一个都行。可惜太阳却并非普通物体，您无法直接拿一把尺子去测量它的大小。当然，您同样也无法直接拿一把尺子去测量它的远近。但您知道，测量物体的远近有一种很常用的间接方法，那就是通过从两个不同的观测点来观测物体，然后利用观测到的角度差异——即所谓的视角差异——来推算它的远近。这种被称为三角视差法（triangulation）的方法从古至今都是测量远近的重要手段（图2.1）。事实上，远在其基本原理被理解之前，我们的大脑和眼睛就已在本能地采用这种方法了，我们的大脑正是利用了左右两眼之间的视角差异，来判断物体远近的。[\[1\]](#)

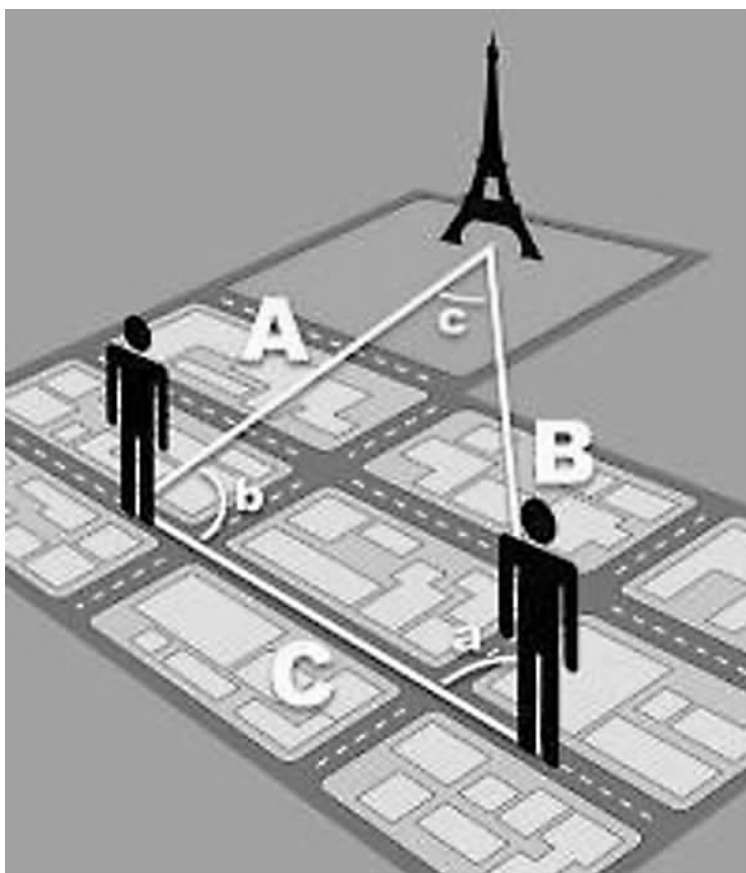


图2.1 三角视差法

但当您试图用三角视差法来测量太阳的远近时，却遇到了巨大的麻烦。三角视差法需要两个观测点，但您很快就发现，从您能够走得到的相距无论多远的两个观测点去看太阳，那视角差异都太小了。地球表面的弧度，地形的细微起伏，乃至您的观测误差都远比您要测量的视角差异大得多。在这种情况下进行测量，犹如在惊涛拍岸声中去倾听远处一只水龙头的滴水之声，您就算长一对兔耳朵也不够用。

怎么办呢？在哪儿才能找到第二个观测点呢？

您冥思苦想了一整夜。当黎明的曙光照到您身上时，您把目光投向了天空。在那里，您看到了一轮淡淡的上弦月（在北半球，上弦月是指右半边可见的“半月”）。看见它，您心中忽然闪过一片灵光，激动得几

乎要像传说中的阿基米德（Archimedes，公元前287—前212）那样一边裸奔，一边大叫：“我找到了！”^[2]

是的，您找到了，您终于找到了第二个观测点，那就是月亮！

别紧张，您没喝酒，您并不是要到月亮上去观测。在古希腊时代人们就已知道，月亮的月相变化并不是月亮本身在变（在古希腊人眼里，天上的东西是永恒不变的），而只是因为阳光从不同角度照射月亮所致。在刚才看见月亮的一刹那，您忽然想到，既然月相是阳光从不同角度照射月亮所致，那它实际上是在告诉您阳光照射月亮的方向，从而也就是太阳相对于月亮的方向。利用这一点，您无需登上月亮就可以推算出从月亮上看太阳的角度，这等于是为您提供了第二个观测点。

特别是，当您看到的月亮恰好是上弦月时，您的视线方向与阳光照射月亮的方向正好是垂直的（图2.2）。这时候如果您记录下太阳的方向，那么它与月亮方向的夹角的一边是月亮到地球的距离，另一边则是太阳到地球的距离，而它的一个邻角恰好是直角。这样简单的三角关系对于即将跻身古希腊先贤行列的您来说无疑是小菜一碟，那两个距离的比值就等于那个夹角的余弦值（cosine）。事实上您还知道，那个夹角的余弦值不仅给出那两个距离的比值，而且还给出了月亮直径与太阳直径的比值。之所以如此，是因为在太阳和月亮之间存在一个美妙的巧合，那就是它们看起来几乎是一样大的。^[3]对于两个看起来一样大的天体，它们与我们距离的比值显然就等于它们直径的比值。

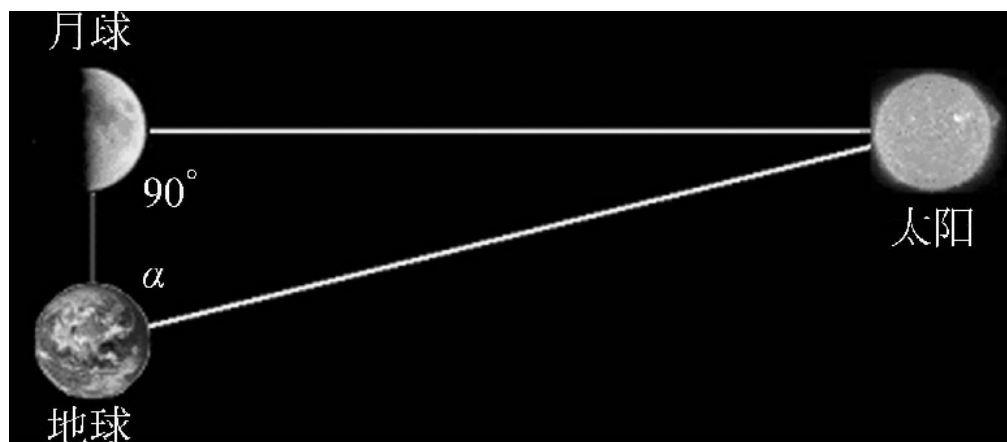


图2.2 太阳、地球与上弦月的相对方位

看来那个夹角很重要，但它究竟是多少呢？那就得靠观测了。不幸的是，那是一个难度很大的观测，因为那个夹角非常接近 90° ，接近到了让您无法分辨的程度。而且在那个夹角如此接近 90° 的情况下，一些在古希腊时代不为人知的因素，比如地球大气对阳光的折射，将足以对结果造成不可忽视的干扰。（感兴趣的读者请想一想，那种影响会使观测到的太阳距离偏大还是偏小？）但不管怎么说，您的方法是正确的，并且即便在当时也有一定的可行性。如果现代人用您的方法来做观测并扣除干扰的话，将会发现那个夹角在 $89^\circ 5' \sim 89^\circ 52'$ 之间。由此得出的结论将是太阳的直径约为月亮直径的400倍，或者等价地，太阳与我们的距离约为月亮与我们距离的400倍。^[4]

这个结果无疑是漂亮的，但与您所要的答案仍有差距，因为它只是把有关太阳的数字和有关月亮的数字联系在了一起，除非您有办法知道有关月亮的数字，它并不能提供您所要的答案。那么，您有办法知道有关月亮的数字，即月亮的大小或月亮离我们的远近（这两个数字您也只要知道其中一个就行了）吗？答案是肯定的。

在常年的天文观测中，您和其他古希腊先贤们一样，已经知道月食

是由于地球挡住了射向月亮的太阳光所致。您并且还注意到，当地球的影子——确切地说是本影（umbra），即完全阻隔阳光的那部分影子——“蚕食”月亮时，影子的边缘是圆弧状的（这是最早使人推测地球为球形的现象之一）。您很快就想到，通过对比影子边缘的形状与月亮本身的形状，您就可以估计出地球影子与月亮的相对大小。不过，这种方法实践起来并不容易，因为地球的影子投射在球状的月亮上并不是一个很简单的几何问题。您想到的一个更好的方法，是对月亮进入地球影子与它穿过地球影子所花的时间进行比较。在前一段时间里，月亮移动的距离等于它自己的直径，在后一段时间里，它移动的距离等于地球影子的直径。因此这两个时间的比值就等于月亮与地球影子的直径之比（当然，这种办法必须要在月亮恰好从地球影子正中间穿过的那种特殊的月食下才能得到可靠的结果）。

如果您进行了那样的测量，您也许会得到一个很接近正确的结果，即地球影子的直径约为月亮直径的2.66倍。^[5]当然，这个地球影子的直径是指地球影子在月亮轨道附近的直径，它要比地球本身的直径来得小（图2.3）。到底小多少呢？几乎恰好小了相当于一个月亮直径的大小（这个结果不是偶然的，感兴趣的读者可以结合太阳比地球大得多，以及太阳和月亮看起来几乎一样大这两点来自行证明一下）。把这个因素考虑在内，您就得到了另一个重要结果：地球的直径约为月亮直径的3.66倍。

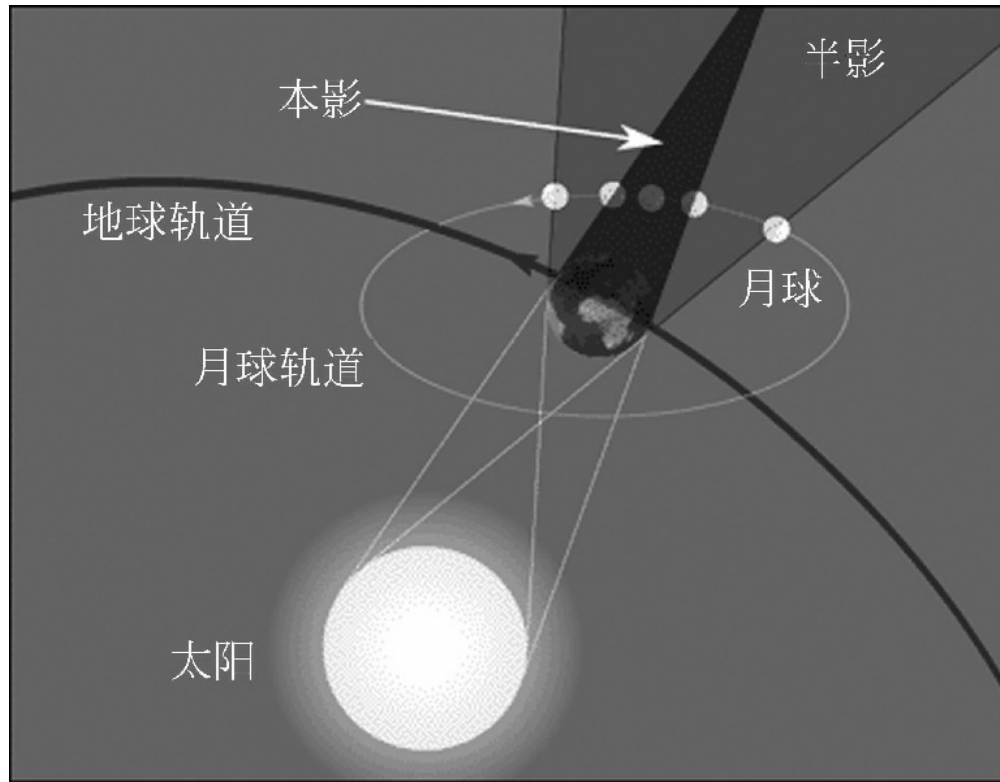


图2.3 测定月亮与地球的相对大小

将这个结果与前面的结果联系在一起，您就发现了太阳的直径约为地球直径的109倍。这个结果意味着太阳是一个庞然大物，在它肚子里可以装下130万个地球——顺便说一下，这是指剁碎了装。如果要问最多能装多少个完整的地球，那可变成一道著名的数学难题了。[\[6\]](#)

就像接力一样，您先把有关太阳的数字与有关月亮的数字联系起来，现在又进一步将它与地球的直径挂上了钩。凭借几何与推理的力量，一个天文问题已被您转变成了地理问题。但问题是，地球虽然就踩在您的脚底下，它的直径却仍然不是可以拿尺去测量的。事实上，在古希腊时代，多数人一生的活动都局限在几千米的范围内，对他们来说，地平线以外的东西就像天边一样遥远。更不用说地球表面的大部分地区被当时还从未有人探索过的汪洋大海所覆盖。

不过您当然不是普通的古希腊人，您总是有办法的。

在所有使人推测地球是球形的天文现象中，除了前面提到的月食时地球影子的边缘形状为弧形外，还有一个很重要的现象，那就是不同纬度的人看到的星空是不一样的。具体地说，那些熟悉的星星或星座在不同的纬度上看时，与天顶的夹角是不一样的。（请读者想一想，为什么我们只提纬度而不提经度？）不仅星星如此，太阳也一样。住在北回归线附近的人大都知道，盛夏正午的太阳是位于天顶正中央的（证据是阳光能直射到垂直深井的底部），而住在北回归线以北的您却发现盛夏正午的太阳是在天顶偏南方向的，具体偏南的角度可以用一根立在地上的垂直杆的投影来计算。这个角度占整个圆周的比例显然就等于您与北回归线的距离（这对您来说是可以测量的）占整个地球周长的比例。由此您就可以计算出地球的周长和直径。经过这样的测量和计算，您发现地球的直径约为12 740千米（当然，这是改用后世的距离单位来表示了）。[\[7\]](#)

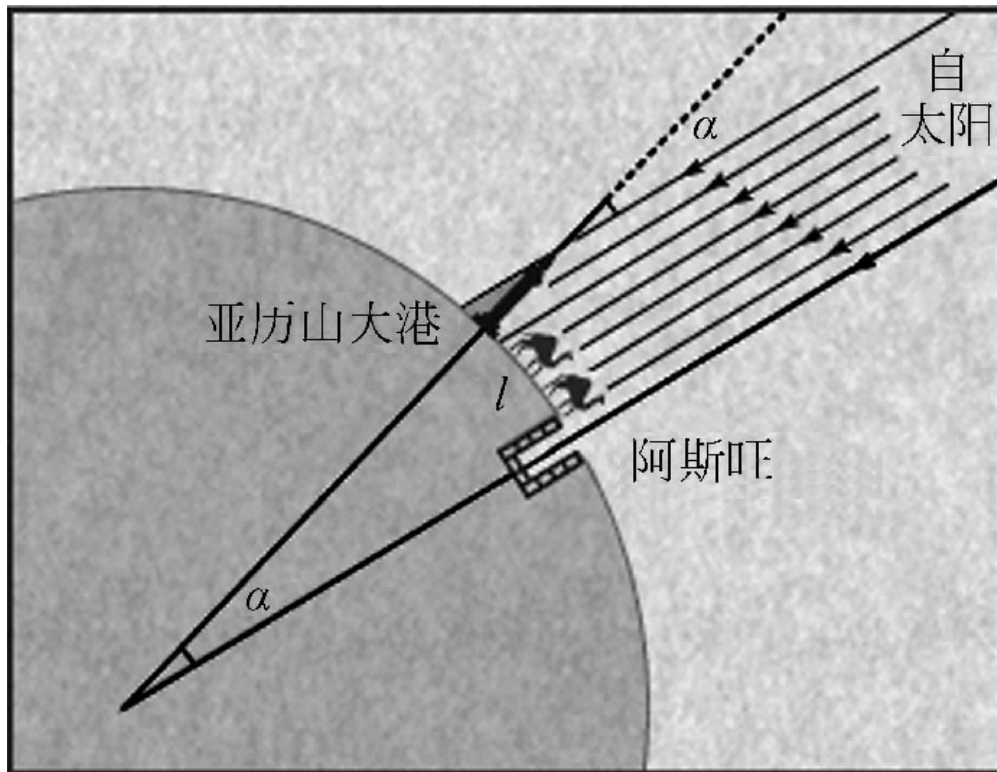


图2.4 测量地球的大小

这样，您就完成了一个漂亮的“三步走战略”：先从太阳到月亮，再从月亮到地球，最后归结到地面上的两个地点，步步相连，环环相扣。将这些环节联系在一起，您就得到了有关太阳的第一个数字：太阳的直径约为139万千米。由此您当然也可以推算出另一个数字：太阳离地球约有1.5亿千米（感兴趣的读者可以脚注提供的数据自行推算一下）。如果您愿意，您还可以写下有关月亮的两个数字：月亮的直径约为3500千米，它离地球约为38万千米（更精确的数字是384 400千米）。[\[8\]](#)

站在我们这个小小星球上，居然能推算出如此遥远天体的性质，这是一件奇妙的事情。在本书的其他章节中，在后来的科学发展史上，这样奇妙的事情还将一再发生。事实上，直到今天为止，除少数飞往过月球，或在近地轨道上生活过的宇航员外，几乎所有人的足迹都从未离开过我们这个小小的星球（包括大气层），但我们却对越来越广阔的外部

世界有了越来越精密的了解。这种能力就是智慧。当然，我们在这里替您稍稍粉饰了一下，限于当时的观测条件，您在数值上是不可能得到像上面那样接近正确的结果的。但对于那个时代来说，最重要的不是数值，而是方法，那一系列精巧的方法足以使您当之无愧地跻身于人类最伟大的先贤之列，永载史册。

您的古希腊虚拟人生兼自助游到这里就结束了，但我们的太阳故事才刚刚开始。接下来，我们将追随历史的足印去探究另外一些重要问题：比如那个肚子里能装下一百多万个地球的庞然大物究竟是什么？它真的是在围绕小小的地球转动吗？再往后，我们还将一起去探究许许多多更现代、更奇妙当然也更困难的问题。

[1] 不仅人类如此，就连某些无法直接利用双眼视差的动物，比如鸽子，也会通过移动自己的脑袋来造成不同的视角，进而判断物体的远近（鸽子虽有两只眼睛，但视野并不重叠，从而不能像人类一样直接利用两只眼睛的视角差异，而只能采用移动脑袋这样的“下策”）。

[2] 传说阿基米德受国王所托，要鉴定一顶皇冠是否掺了杂物。他苦思良久，最终在洗澡时悟出了用浮力进行鉴定的方法，欣喜若狂的他连衣服都没穿就冲出浴室大喊：“我找到了！”这个故事并未被记录在阿基米德著作之中，它的真实性后来引起了一些有趣的争议。

[3] 确切地讲，由于地球绕太阳和月亮绕地球的公转轨道都是椭圆，太阳和月亮看起来的大小都不是不变的。其中太阳的角直径最小时为 $31' 27.7''$ ，最大时为 $32' 31.9''$ ，平均为 $31' 59.3''$ ；月亮的角直径最小时为 $29' 23.0''$ ，最大时为 $33' 31.8''$ ，平均为 $31' 5.3''$ 。我们在后文中将会看到，月亮的角直径有时比太阳大，有时又比太阳小这

一特点对于日食的种类有着很重要的影响。

[4]在历史上，古希腊先贤阿里斯塔克斯（Aristarchus，公元前310—前230）曾经用这种方法进行过测量。他估计出那个夹角为 87° ，与实际数值只相差不到 3° 。可惜对于这种很接近 90° 的角度来说，哪怕只相差 1° 也足以造成很大的误差。阿里斯塔克斯估算出的太阳直径只有月亮直径的18~20倍。

[5]在历史上，这个方法也同样被阿里斯塔克斯采用过，他估计出的地球影子直径约为月亮直径的两倍，由此得到的太阳直径约为地球直径的7倍。这个结果虽然误差极大，但——如我们在下一章中将会提到的——仍给了阿里斯塔克斯一个很重要并且很正确的启示。阿里斯塔克斯之后的其他先贤们对地球影子的直径给出了更好的估计，比如希帕克（Hipparchus，公元前190—前120）给出的估计为月亮直径的2.5倍；托勒密（Ptolemy，90—168）给出的估计为月亮直径的2.6倍。

[6]这个数学问题被称为“开普勒猜想”（Kepler Conjecture），是一个著名的数学难题。1998年，美国数学家黑尔斯（Thomas Hales，1958— ）发表了一个长达250页，并且需要计算机辅助的证明，但该证明迄今尚未得到数学界的公认。

[7]在历史上，古希腊先贤埃拉托斯特尼（Eratosthenes，公元前276—前195）曾经用这种方法估算过地球的周长。图2.4中那两个城市（即阿斯旺和亚历山大港，纬度分别为 $24^\circ 05' \text{ N}$ 和 $31^\circ 02' \text{ N}$ ）就是埃拉托斯特尼所选的观测点。由于史学界对埃拉托斯特尼所用的距离单位尚有争议，今天我们尚无法确切知道他的估算结果，但一般认为是在39 690~46 620千米之间（相应的直径在12 630~14 840千米之间）。

[\[8\]](#)有意思的是，月亮的存在对于上述推理具有极大的重要性。事实上，如果没有月亮，人类科学的很多早期探索都会遇到额外的困难。



绘画：张京

3 地心说vs日心说

我们已经知道，天上的日月星辰并不是静止不动的，从它们的东升西落中所能得到的最直接、最直观的结论，就是所有天体都在一个以地球为中心的天球上，围绕地球转动。这种几乎出现在所有早期文明中的猜测是地心说（Geocentrism）的雏形。

但世界的有趣之处就在于，它常常给你一点希望，似乎一个简单图像就能让你抓住点什么，但稍稍细究一下会发现事情并不那么妥帖。拿日月星辰的运动来说，星星的运动倒是的确能用一个天球的转动来描述——因为它们只有周日运动，但太阳、月亮及五大行星却除了周日运动外还各有各的“私活”：太阳有周年运动，月亮有月相变化，五大行星更不像话，不仅各有各的周期，甚至还每隔一段时间就“倒行逆施”（逆行）一番。区区一个天球是无论如何也摆不平那么多运动的。怎么办呢？古人们想到了一招，那就是把天球当成礼物派发，让太阳、月亮及五大行星各占一个，乖乖听话的其他星星们则共享一个。^[1]

但这还不够，因为行星的逆行还无法解释。有人也许会说，那有什么难的？让天球一会儿正转，一会儿逆转不就行了？打住！万万不行。要知道，从古希腊开始直到17世纪之前，在差不多两千年的时间里，人们对天体运动的描述一直遵守着两个要素：一是天球必须为球形，二是它的运动必须有某种类型的均匀性。这几乎是当时对“解释”一词的定义，非如此不能算是解释。让天球像眼珠子那样乱转是万万不行的——文雅点说是不完美的。

天球必须完美，行星却要倒行逆施，这就让人伤脑筋了。在被伤了

脑筋的人当中就有古希腊先贤柏拉图（Plato，公元前428（427）—前348（347）？），他给后人留了一道思考题：如何用均匀有序的运动来描述看起来不规则的行星运动？

要说历史上的聪明人还真不少，柏拉图的思考题一出，很快就有人按下了抢答键。抢答者不是外人，而是柏拉图的学生欧多克斯

（Eudoxus，公元前400（408）—前355（347））。他的答案很豪爽，那就是派发更多的天球，让每个行星都被几个同心天球共同带动，直到满意为止。这个答案的妙处在于可以让那些同心天球的转轴及快慢彼此不同，但却各自保持均匀（从而仍然是完美的）。在欧多克斯的模型中一共用到了27个天球，用这种方法，他对包括行星逆行在内的很多天体运动现象给予了粗略描述。为了表彰他的贡献，我们授予他一个荣誉称号：第一位试图对行星运动做出数学描述的先贤。

但欧多克斯的模型无法经受住哪怕只是稍微细致一点的观测考验，而且人们早就发现行星的亮度并不是恒定的，在当时这意味着它们与地球的距离不是恒定的，这显然不是欧多克斯的同心天球模型所能解释的。怎么办呢？另一位古希腊先贤，以研究圆锥曲线著称的阿波罗尼斯（Apollonius，公元前262—前190）支了一个妙招。阿波罗尼斯提出太阳、月亮及五大行星各自绕一个所谓的本轮（epicycle）作匀速圆周运动，而本轮的中心则绕一个以地球为中心的所谓均轮（deferent）作匀速圆周运动。用这种方式，他不仅可以让行星与地球的距离发生变化，而且同样可以——并且能更好地——解释行星的逆行。不过在他的模型中出现了不以地球为中心的东西——本轮，这对最刻板的地心说模型是一种偏离。这种偏离是在纯粹观念与观测现实之间小心翼翼的妥协，它看似细微，却是一种实证精神的萌芽。

阿波罗尼斯的这种均轮加本轮的构想成为了地心说模型的新框架。

为了拟合越来越精密的观测，地心说模型变得越来越“轮”丁兴旺。但不幸的是，人们很快就发现，轮多不一定力量大，有一些细微现象，比如行星的逆行幅度时大时小，似乎无法靠简单地增添轮子来解释。怎么办呢？地心说模型的集大成者，古希腊天文学家托勒密（Ptolemy，90—168）决定下猛药，让纯粹观念再次向观测现实作出妥协——而且是重大妥协。托勒密一举放弃了均轮的中心为地球，以及均轮的转动为匀速这两大几乎被视为底线的观念，引进了所谓偏心等距点（equant point）的概念，对诸如行星逆行幅度时大时小之类的现象作出了一定程度的解释。[\[2\]](#)



古希腊天文学家托勒密（90—168）

但这还不够，因为除上述现象外，行星运动还有一个引人注目之

处，那就是水星和金星的运动总是局限在太阳左右一个小范围之内，而不像其他行星那样满世界乱跑。为了解释这一现象，这两颗行星的本轮中心被假定为永远处在地球与太阳的连线上。把这些修正汇集到一起，托勒密的地心说模型（彩图3）就基本完成了，剩下的只是微调。

后人对托勒密这个与地心说联系在一起的名字往往有一种模糊的错觉，以为那是某种保守理论的代言人。事实上，托勒密是一位真正伟大的天文学家，他的伟大不仅体现在他自己的时代，甚至还向后延伸了一千多年。在那个重理念轻实践的时代里，他明确提出理论必须拟合观测。他的地心说模型就是这种努力的典范，其精度之高，甚至连一千四百年后的日心说模型也无法轻易超越。当代科学史学家霍尔顿（Gerald Holton, 1922— ）曾有过这样一句感慨：“没有什么事情比低估古希腊人的观点更容易和更错误。”我有同感。当我们挟两千年的历史优势来回顾某些古希腊先贤的思想时，理所当然地会看到大量的错误，甚至可以不夸张地说，他们写得越多，就错得越多（相比之下，像“道可道，非常道”那样东方式的言简意赅实在是太有才了）。但那些错误就像孩子学步时的摔跤，没有它们，人类恐怕直到今天还在爬。

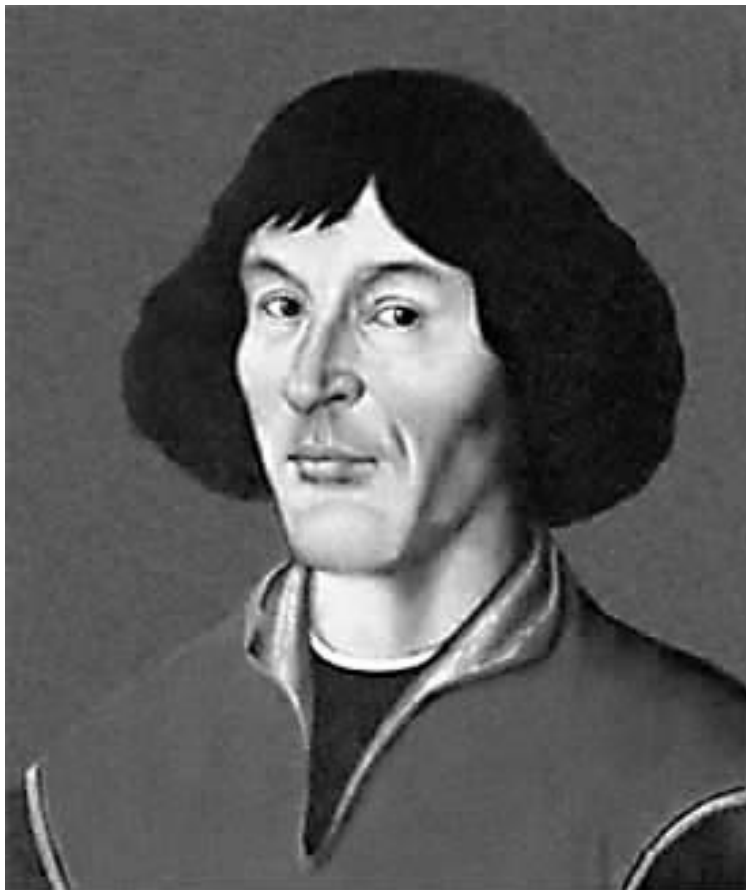
在接下来一千多年的时间里，地心说模型作为天体运动的主流模型，成为了导航、测绘、及星图计算的基础。不过，在这一模型的发展过程中，如上所述，实证精神已开始萌发。随着偏心等距点等概念的提出，诸如天球以地球为中心以及天球必须作均匀圆周运动那样的古老信念已被显著削弱。而且地心说虽然解释了不少现象，带来的问题却也不少。对那些问题的探究使一些人进行了不同的尝试，其中最早的努力甚至在托勒密之前就出现了。

我们在上一章的注释中提到过，古希腊先贤阿里斯塔克斯（Aristarchus，公元前310—前230）曾经用我们介绍过的方法估计过太

阳的直径，结果约为地球直径的7倍（相应的体积约为地球体积的343倍）。虽然与现代值相差很远，但太阳比地球更大这一定性结果还是给了阿里斯塔克斯很大的启示。要知道，地心说的产生在很大程度上乃是出自直觉，而这直觉有两个来源，一个是天体的周日运动，另一个则是天体看上去都很小——小东西围绕大东西转似乎是天经地义的。但如果太阳比地球更大，这直觉就成问题了。一个庞大的太阳有什么理由要绕一个体积不到自己1/300的小不点儿转呢？一般认为，正是这个怀疑使阿里斯塔克斯提出了太阳才是宇宙中心的观点，这是最早的日心说（Heliocentrism）。^[3]

但阿里斯塔克斯的日心说并未引起什么反响，因为它面临很多棘手的问题，比方说如果地球在运动，那天上的云彩为什么不会被运动的地球所抛离？这个问题别说阿里斯塔克斯，就连一千八百年后的哥白尼（Nicolaus Copernicus, 1473—1543）也难以回答。另一个问题比较有意思，那就是如果地球在运动，那么星星的位置应该会像用三角测距法测距时那样显示出视差（因为观测点随地球运动而变化），但实际上我们却从未发现过那样的视差，这是为什么？对于这个问题，阿里斯塔克斯作出了很正确的回答（可惜没人相信），那就是星星离我们实在太远，以至于视差小到了无法被察觉的程度。他的这个回答本身就是一个了不起的天文发现，因为它给当时尚无人知晓的宇宙大小设置了下限，即宇宙起码要大到能让星星的视差不被肉眼所察觉的程度。第三个问题来自所谓“天贵地贱”的观念，当时的很多人相信天上的星星是永恒而完美的，地上的一切则是腐朽而卑微的，两者无论在外观还是质料上都截然不同。^[4]而日心说却要让腐朽卑微的地球混迹于永恒完美的行星行列，这怎么可以呢？这个问题在今天看来很无厘头，但在当时却是难以抗拒的“主流民意”。

虽然阿里斯塔克斯的日心说未能掀起波澜，但地心说的麻烦却并未结束。除了太阳比地球大所导致的困扰外，地心说还有其他一些不如人意之处。比方说行星和太阳在地心说中是有相似地位的，但行星的本轮周期却全都是一年，即恰好等于太阳绕地球运动的周期，这种巧合在地心说中是很难解释的。此外，随着航海业的兴起及对日历与定位精度的要求日益提高，地心说的精度也越来越成问题了。正是在这种背景下，1543年，一本全面阐述日心说的著作——《天体运行论》（*On the Revolutions of the Heavenly Spheres*）问世了。这是一部“难产”的著作，它的作者——波兰天文学家哥白尼——用了长达23年的时间来撰写它，完成之后又因担心触怒教会（同时也为了进行细节完善）而延迟了13年，直到去世前不久才发表。



波兰天文学家哥白尼（1473—1543）

哥白尼的这部著作是自托勒密以来最杰出的天文学著作，哥白尼虽然不是最早提出日心说的人，却是最早将日心说由一个观念性学说转变为具有预言能力的定量模型的人。^[5]在哥白尼的日心说模型中，我们这个系列的主角——太阳——荣升为了宇宙的中心，我们脚下的地球则变成了行星，一边自转，一边和其他行星一样围绕太阳公转（图3.1）。地心说无法解释的行星本轮周期全都是一年的巧合在日心说中变得显而易见，因为那不过是地球公转产生的表观现象。天体的周日运动也有了很简单的解释，即那不过是地球自转产生的表观现象。不过哥白尼的日心说模型在最低阶近似上虽比同等近似的地心说模型高明得多（因为无需引进本轮），但由于和地心说模型一样未能摆脱圆周运动这一束缚（在这方面哥白尼甚至比托勒密更保守，连偏心等距点那样的概念都不曾引进），从而一涉及细微之处，就无论以繁简程度还是精度而论，都无法真正超越托勒密的地心说模型，这一点直到17世纪初德国天文学家开普勒（Johannes Kepler, 1571—1630）发现椭圆轨道后才得以改变。

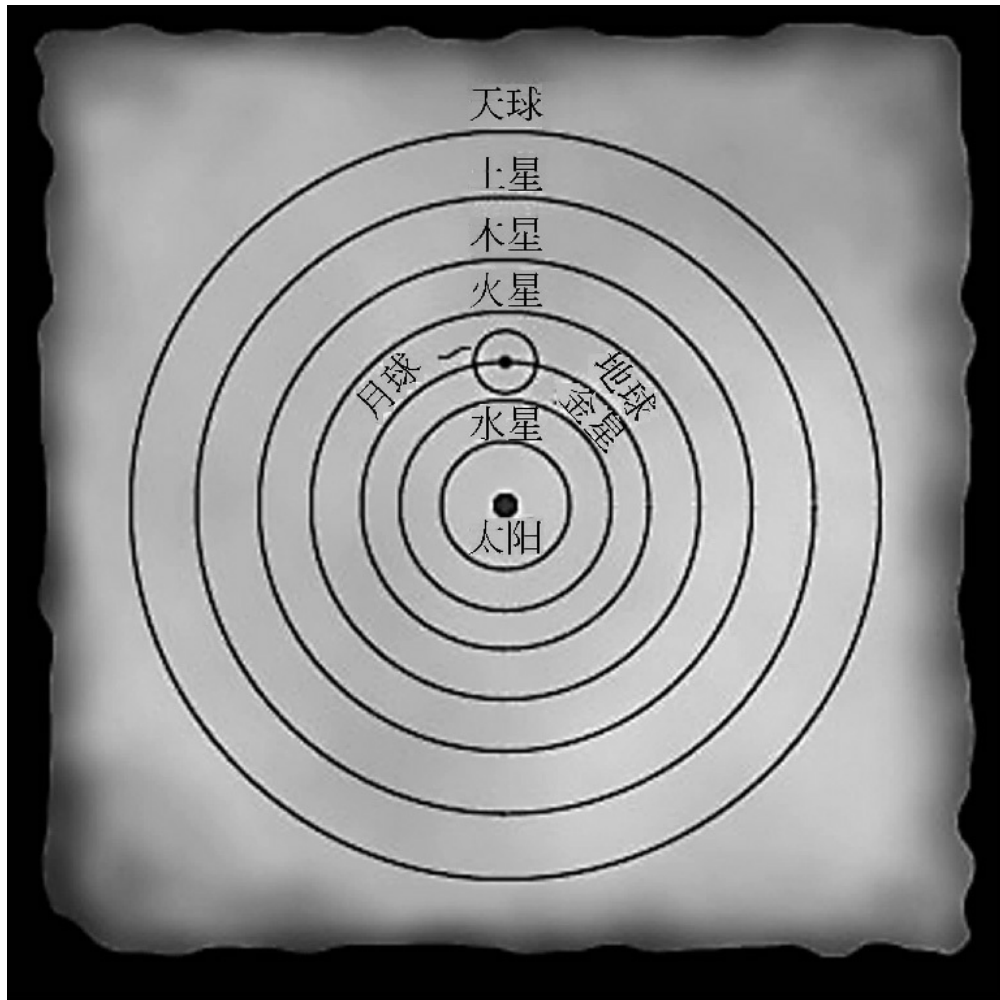


图3.1 简化版的日心说模型

我们刚才提到，哥白尼曾经担心自己的日心说会触怒教会。这种担心并非杞人忧天，因为《圣经》中有不止一处提到太阳运动而地球静止，日心说与那些文字是有冲突的。不过具有讽刺意味的是，哥白尼时代的教会虽竭力维护托勒密的地心说，仿佛后者是天经地义的真理，但实际上，无论托勒密的地心说，还是资格更老的亚里士多德

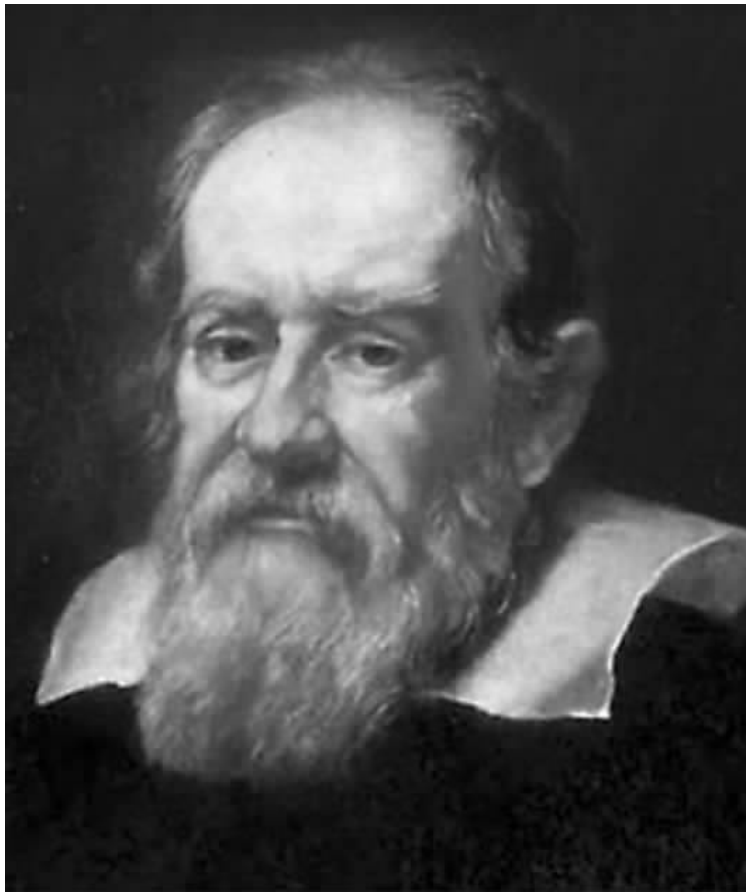
（Aristotle，公元前384—前322）的著作，都并非一直就是教会的宠儿。相反，它们都曾经上过教会的黑名单——一度被教会所禁止。因为那些著作的逻辑与实证色彩和教会所希望的盲从与盲信背道而驰，而且它们在文字上虽对上帝充满了虔敬，其所宣扬的世界体系却基本无需上

帝的帮助，把上帝这个“活雷锋”架空了。不过在13世纪中期，教会采取了新的策略，对一些有影响力的自然哲学著作进行“无菌处理”，使之与圣经接轨。这样做既减少了对手，又充实了自己的理论阵地，可谓一举两得。^[6]在哥白尼时代受教会维护的托勒密地心说就是这种经过“无菌处理”后具有教会特色的自然哲学。而哥白尼要推出一种“带菌”的学说，心中自然不无忐忑。

不过忐忑归忐忑，哥白尼与教会的关系其实还是蛮“和谐”的，因为他的正业其实是神职人员（因此有人戏称哥白尼白天是神父，晚上才是天文学家），他甚至把自己的《天体运行论》献给当时的教宗保罗三世（Pope Paul III, 1468—1549）。而《天体运行论》的出版者奥希亚德（Andreas Osiander, 1498—1552）更是为该书包上了一层精心制备的“糖衣”——序言，谦虚地宣称日心说仅仅是一个便于计算的假设，不一定是真实的。经过这样“瞒天过海”的包装，《天体运行论》这一“糖衣炮弹”的发表并未受到教会的干预。直到几十年后意大利科学巨匠伽利略（Galileo Galilei, 1564—1642）支持日心说的著作引起强烈反响后，教会才意识到自己被忽悠了。

伽利略的工作为日心说的胜出提供了强有力的支持。我们前面提到过，在地心说模型中水星和金星的本轮中心被固定在地球与太阳的连线上（图3.1），按照那种模型，相对于地球来说，太阳永远只能从后方或侧后方照射水星和金星，由此导致的结果是这两者永远只能有弯钩状的相（感兴趣的读者请自行证明这一点）。但伽利略却通过望远镜发现金星具有类似“满月”那样的相，从而对地心说构成了判决性的打击。^[7]另一方面，在所有针对日心说的反对意见中，除宗教因素外，最棘手的问题是运动地球上的东西为什么不会被地球所抛离？这个问题难倒了从阿里斯塔克斯到哥白尼的所有人，却被伽利略提出的相对性原理所回

答。甚至连“天贵地贱”那样的“主流民意”，也因伽利略发现太阳黑子而遭到了驳斥。



意大利科学巨匠伽利略（1564—1642）

1632年，伽利略发表了名著《关于两大世界体系的对话》（*Dialogue Concerning the Two Chief World Systems*，简称《对话》），对地心说和日心说这两大世界体系进行了看似不偏不倚，实则显著青睐后者的阐述，而且在阐述中还对已被教会认可的亚里士多德的很多观点提出了异议。这一切令教会很生气，后果很严重。其实，伽利略倒也并非吃了熊心豹子胆，他在1616年曾受到过教会的“警告处分”，在那之后安静过几年。这回之所以敢“顶风作案”，是因为他的一位大主教朋友于1623年成为了教宗厄本八世（Pope Urban VIII, 1568-1644）。这么过硬

的上层关系给了伽利略一种安全感，使他以为“科学的春天”到来了。

他完全错判了形势。

实际情况是：《对话》发表后才不过一年，1633年，罗马宗教裁判所就对伽利略进行了审判，并裁定他有罪。低头认罪还是顽抗到底？这是一个问题。年近古稀的伽利略选择了前者，他在认罪书中表示：“我，伽利略，……手按圣经起誓，我过去和现在一直相信，在主的帮助下今后也将相信圣天主教和使徒教会所持有、传授及教导的一切。”对于日心说，他表示：“我发誓今后绝不以口头或书面形式发表任何类似的东西。”

但教会的干预最终未能阻止科学在经历了中世纪黑暗后的快速复兴。日心说先是作为一种纯粹的计算工具（如《天体运行论》的序言所谦称的那样），而后作为一种具有真理性的理论还是逐渐流行了起来。1832年，苏格兰天文学家亨德森（Thomas Henderson, 1798—1844）发现了半人马座 α 星（ α Centauri）的视差，[\[8\]](#)从而直接证实了阿里斯塔克斯的先见之明，同时也扫清了日心说的最后一个技术障碍。

在无可逆转的局势面前，教会的态度也终于有了变化。1992年10月，教宗约翰·保罗二世（Pope John Paul II, 1920—2005）向伽利略在引进实验方法及理解日心说方面所做的巨大贡献表示了敬意和感谢，并承认教会对地心说的维护乃是对《圣经》作字面解读导致的错误。2000年3月，约翰·保罗二世签署正式声明，承认当年对伽利略的审判是错误的，并为教会两千年来所犯的暴力、迫害及错误道歉。2008年3月，梵蒂冈教廷为伽利略建造了雕像——离那雕像的矗立之处不远，便是375年前伽利略等待审判的地方。

“无可奈何花落去，似曾相识燕归来”，历史走过了一个漫长而沉重的轮回。

日心说与地心说的争论虽然以日心说的胜出而落幕，但日心说将太阳视为宇宙中心的做法却很快也受到了质疑。事实上，太阳与星星的差别不就是前者看上去像一个圆盘而后者像一群小点吗？既然日心说已经把我们脚下这个看起来比太阳还大的地球与那五个看上去像小点的行星归入了同一类别，太阳为什么就不能和星星一样呢？

思想的禁忌一旦被打破，新的想法就会源源而生。1584年，意大利哲学家布鲁诺（Giordano Bruno, 1548—1600）提出了太阳只是无限宇宙里的无数颗星星之一的想法。^[9]1644年，法国数学家笛卡儿（René Descartes, 1596—1650）也提出了太阳和其他星星一样的观点（虽然他有关太阳和星星的具体模型——旋涡模型——是完全错误的）。今天我们知道，太阳只是太阳系的中心，而不是宇宙的中心。太阳是银河系上千亿颗恒星中很普通的一员，而银河系又只是可观测宇宙上千亿个星系中很普通的一员。

这就是太阳的真正身份——一颗非常普通的恒星，可观测宇宙中发光星体的一百万亿亿分之一。

但无论多么普通，太阳对于人类来说都是独一无二的。它是光和热的源泉，也是唯一一颗能让我们看到圆面的恒星。它与我们的关系是如此密切，在任何一个晴朗的白天，它那光芒夺目的圆面哪怕只出现些许的破缺，或几分钟的消失，都是非同小可的事情。在下一章里，我们将一同去欣赏这种被称为日食的现象。

[1]顺便说一下，从那时起直到中世纪，太阳和月亮是被很多人视为行星的（因此那时的行星实际上有7颗，太阳和月亮都曾拥有一个名字叫行星），我们为了避免混淆才将它们单独列出。

[2]托勒密的偏心等距点概念是对他所放弃的以地球为均轮中心，以及均轮的转动必须均匀这两大观念的补偿。在他的模型中，地球虽不再处于均轮的中心，但和一个被称为偏心等距点的位置对称分布在均轮中心的两侧，均轮转动的角速度相对于球心和地球虽不再均匀，但相对于偏心等距点却是均匀的。这个模型可以说是天球体系中最接近椭圆轨道的构想（偏心等距点的作用接近于椭圆的焦点）。

[3]阿里斯塔克斯是最早提出地球绕太阳运动的人，但在他之前就有人提出过地球运动的想法，比如毕达哥拉斯学派的古希腊先贤菲洛劳斯（Philolaus，公元前470—前385）曾提出过包括地球、太阳在内的所有天体都围绕一个所谓的“中央火”（central fire）转动的观点。

[4]细心的读者也许会问：我们用肉眼就能看到月亮上的阴影，这岂不是与天体的完美相矛盾？Good question！古希腊人对此的解释是月亮离地球太近，从而被“污染”了（近墨者黑嘛）。有人可能还会进一步问：为什么要让一大批永恒完美的天体围绕一个腐朽卑微的地球转呢？这不是把鲜花插在牛粪上吗？这个大逆不道的问题就得问上帝了，据说是它把人类造在地球上，并让天体围绕地球转的。

[5]在从阿里斯塔克斯到哥白尼的一千八百年间，有一些其他人也提出过背离地心说或接近日心说的理论，但都没有定量化，且大都因不符合“主流民意”而自行或被迫放弃了。

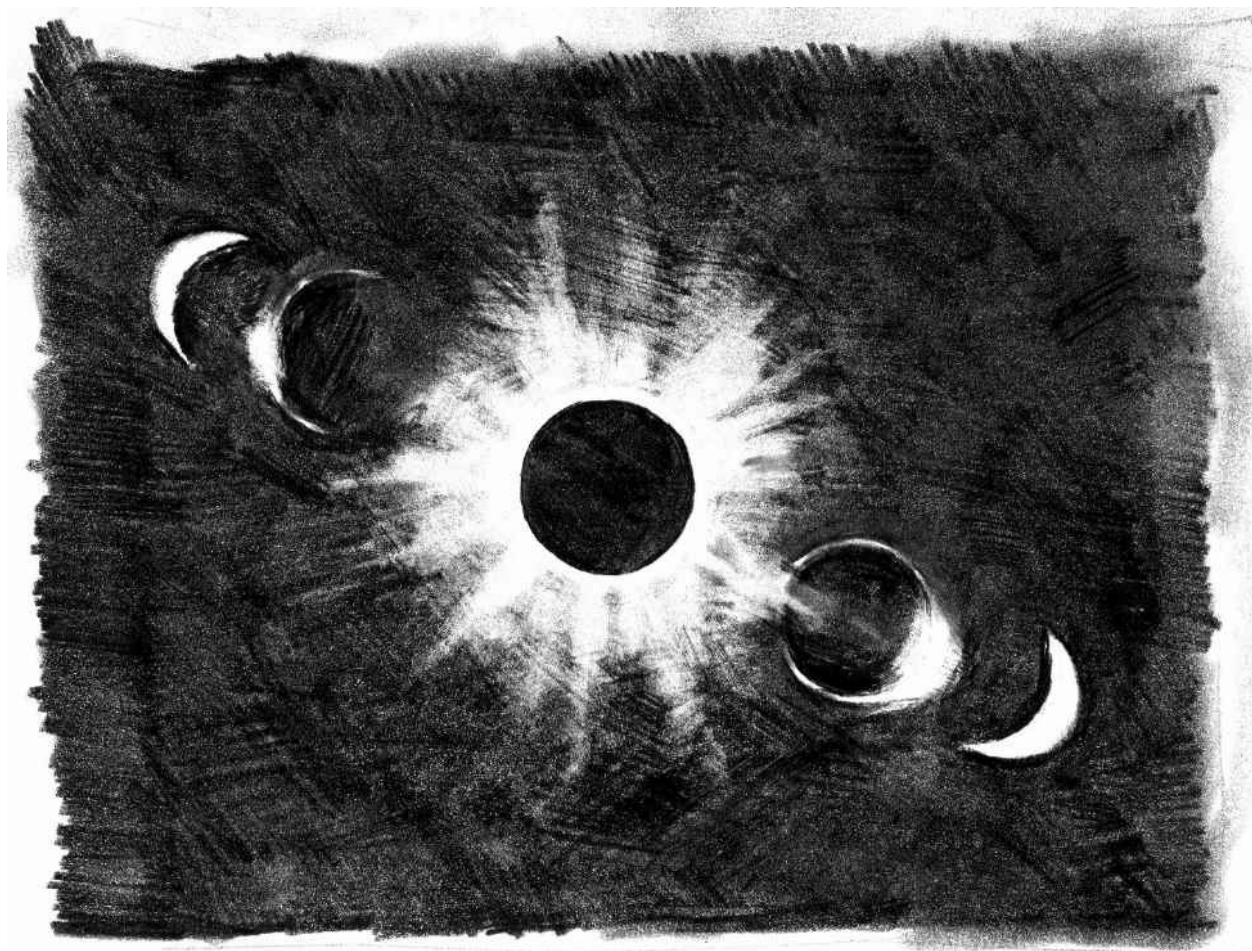
[6]在这一过程中为教会立下汗马功劳的是所谓的经院哲学家

(scholastic philosopher)，其代表人物是被教会封为“天使圣师”的阿奎纳斯(Thomas Aquinas, 1225—1274)。经院哲学家以宗教教义为指导，通过对古典哲学思想进行包装、诠释及歪曲、诡辩，而构建出与圣经相协调的新体系。

[7]在日心说模型中，金星的“满月”相出现在它与地球分处于太阳两侧的情形下，这时金星离地球的距离最大，从而部分抵消了“满月”相的亮度优势，因此金星的亮度变化远小于单纯由距离变化所预期的，这一点曾被认为是日心说的一个难题。

[8]不过亨德森对自己的观测缺乏信心，直到1839年1月才发表，在发表时间上落后于德国天文学家贝塞尔(Friedrich Bessel, 1784—1846)，后者于1838年发现了天鹅座61星(61 Cygni)的视差。

[9]布鲁诺因为这一信念及其他若干罪名被罗马宗教裁判所起诉，并最终判处火刑，于1600年2月17日处决。



绘画：张京

4 日食——既寻常又稀有的奇观

美国科幻作家阿西莫夫（Isaac Asimov, 1920—1992）曾经写过一个著名的科幻故事，叫做《黄昏》（*Night fall*）。在这个被许多读者推举为有史以来最优秀科幻作品之一的名篇中，阿西莫夫虚构了一个由六个太阳组成的多星系统，在那里有一个有“人”栖居的行星上，几乎任何时候都至少有一个太阳悬在空中，夜幕每隔2049年才会降临一次。每当那一时刻来临，地平线上硕果仅存的一个太阳会遭遇“日全食”（total eclipse），传说中能夺人魂魄、让人丧失理智，进而毁灭整个文明世界的星星会出现在黑暗天空里。所有人都在短时间内陷入巨大的恐慌和骚乱之中，周而复始地亲手将星星毁灭文明世界的传说变为现实。

我们很幸运，没有生活在那样的多星系统中。^[1]只有一个太阳的我们早已习惯了日夜的更替，星星在我们眼里非但不是毁灭文明世界的恐怖象征，反而是一种魅力无穷的美景（从这个意义上讲，它倒也能夺人魂魄）。不过，即便对于我们来说，如果太阳在原本应该出现的时候突然消失，或其圆面出现破缺，依然是一件非同小可的事情。在这种被称为日食（solar eclipse）的自然现象中，最扣人心弦就是阿西莫夫故事中提到的日全食。如彩图4所示，当太阳即将被全部遮盖时，月球的巨大阴影自西向东掠过地面^[2]，夜幕骤然降临，星星重新闪耀，有时还能看见美轮美奂的贝利珠（Baily's bead）和壮丽的日冕（solar corona）。这时候，有些人将会感到颤栗——不是因为害怕，而是由于气温的骤降。在日全食的短暂过程中，气温会有相当明显的降低（比如1878年7月29日的日全食期间，美国丹佛市的气温从酷热的46°C骤降到28°C）。这种笼罩整个天空和大地的大自然的华丽演出，对亲临者来说往往是令人陶

醉和毕生难忘的。

不过这只是现代。在古代，日食的出现可不是什么轻松的事情（虽然有时也能导致好结果）。古希腊历史学家希罗多德（Herodotus，公元前484—前425）曾经记载过公元前六世纪米底斯（Medians）与利底亚（Lydians）两族交战时，因遭遇日食致使双方惊恐万分、终止干戈的事情。那样的事情在历史上发生过多，甚至直到20世纪60年代，柬埔寨内阁首相朗诺（Lon Nol，1913—1985）的军队还在一次战斗中因遭遇日全食而惊慌失措。

在对日食成因缺乏了解的古代，人们对日食有过许多荒诞不经的猜测。我们熟悉的“天狗吃太阳”就是其中之一。为了阻止天狗的“暴行”，民众常常敲锣打鼓、鸣金放炮，甚至连九五之尊的皇帝也往往要吃素斋、避正殿，诚惶诚恐。类似于“天狗吃太阳”的传说在其他古文明也有，只是所涉及的猛兽不尽相同，吞吃太阳的原因五花八门而已。最搞笑的或许是印度的一种传说，该传说认为太阳常常向某个野蛮部落的成员借钱，却借而不还。对方一怒之下就会吃掉它（但吃完后又吐出来，看来还是惦记着还钱的事情）。这种猛兽（或猛人）吃太阳的说法并不是古人对日食成因的唯一猜测。比这种“猛兽说”更有情趣的是所谓的“夫妻说”。那种说法也是形形色色，比如在某日耳曼传说中，月亮和太阳是一对不和睦的夫妻，而且太阳是妻子，当月亮偶尔去看望太阳时就会出现日食，但它们和睦不了几分钟就又会闹翻（幸亏如此），于是太阳就会重新出现。当然，也有些传说不那么“诬蔑”太阳和月亮的夫妻关系，比如在北美某印第安部落的传说中，太阳和月亮是恩爱夫妻，日食的成因——猜得到算你有本事——是太阳的手臂上抱着他们的孩子！

这些荒诞不经的传说，以及视日食为恐怖天象的年代对于生活在现代文明社会中的人来说已经很遥远了。对现代人来说，有幸观赏到日

食，非但没什么恐怖，反而是令人羡慕的经历。有些天文爱好者甚至不惜破费，千里迢迢地赶往日食带观赏日食。有些天文学家更是甘当现代“夸父”，乘坐喷气式飞机追逐日食，目的就是让这一天象的持续时间尽可能延长。我们早已知道，日食并不神秘，它只不过是月球偶尔跑到地球与太阳之间，挡住了阳光而已。

明白了这个浅显的道理，我们就可以冒充天文学家来分析一下日食的规律。我们知道，月球每个“月”都会绕地球转一圈。既然是转一圈，那就总会经过太阳的方向。这么推算起来，日食似乎每个月都应该发生一次，每个世纪则应该有1 238次。（请读者想一想，为什么不是1 200次？）但事实上，在已经过去的20世纪中，总共只发生过228次日食。在21世纪中，也将只会有224次日食，只占1 238次的18%左右。

看来天文学家不是那么容易冒充的，我们首先得搞明白：那另外82%的日食哪里去了？

答案是：被地球漏掉了。要解释这一点，我们需要对地球与月球的公转轨道特点有一个简单了解。这其中对日食来说最重要的特点，是地球与月球的公转轨道并不在同一平面上，而是有一个平均为 5.14° 的夹角。由于这个夹角的存在，即使月球位于地球和太阳之间，它相对于地球公转轨道平面来说，也往往不是偏上，就是偏下，它的影子多数时候只能投射在清冷的虚空之中，而无法触及地球，从而无法形成日食（图4.1中左、右两侧的情形就是如此）。[\[3\]](#)

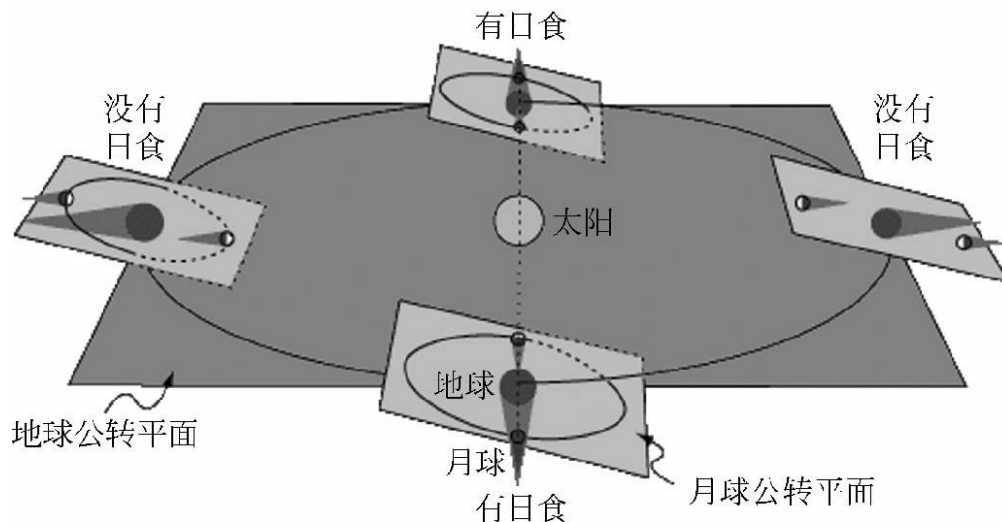


图4.1 地球、月球的公转轨道与日食

那么什么时候才会有日食呢？显然只有当月球不仅位于地球和太阳之间，而且还恰好离地球公转轨道平面不远时。由于整个月球公转轨道（作为一个椭圆）与地球公转轨道平面的交点只有两个，因此月球只有运动到地球和太阳之间，并且又恰好在这两个交点之一的附近时，才能形成日食。在天文学上，这两个交点有自己的名字，视月球自南向北还是自北向南穿越地球公转平面而定，分别被称为升交点（ascending node）和降交点（descending node），两者的连线则被称为交点线（line of nodes）。用这种术语，形成日食的条件也可以表述为：月球运动到地球和太阳之间，并且交点线与太阳方向几乎重合（这保证了月球在两个交点之一的附近）。从图4.1中不难看到，交点线与太阳方向几乎重合的情形在整个地球公转周期上只会出现在两个时段里（即图4.1中两个“有日食”的位置附近），因此日食的发生并非每月一次，而似乎是每年只有两次。

如果每年有两次日食的话，一个世纪就是200次。这与前面提到的次数接近了不少，但仍有偏差，可见在日食的频率中还隐藏了别的玄机。这玄机之一就在于月球的公转轨道并不是固定不变的。主要受来自

太阳的引力摄动影响，月球的公转轨道平面在不断改变着方位，由此导致的结果是交点线会缓慢转动，这种转动的方向与地球公转的方向相反，周期约为18.61年，称为交点的退行（regression of nodes）。如果不存在交点的退行，交点线将每隔半年与太阳方向重合一次，日食将会有规律地出现在每年年历的固定时段中，它在历史上的神秘感也将褪去不少。但由于存在交点的退行，日食的出现时间就变得有些扑朔迷离了。因为交点线离开太阳方向后，无需经过半年，就会由于自己的转动而提前与太阳方向再次重合。简单的计算表明，这之间的间隔只有173.3天（读者可以利用交点线的转动周期为18.61年这一信息自行计算一下）。两次这样的间隔则是346.6天，这称为交点年或食年（eclipse year）。这个时间单位在分析日食时有着重要作用。

由于每个食年都包含了交点线与太阳方向的两次重合，而“食年”的长度（346.6天）只有“年”（365.24天）的95%左右。这么推算起来，每个世纪包含的食年数目约有105个，而日食的次数则应该有210次左右。这比前面的200次又进了一步，但仍比实际次数来得少。这又是怎么回事呢？这次的答案更加微妙，但关键之处其实已蕴涵在了前文的措辞之中。我们在前面提到，日食发生的条件是月球运动到地球和太阳之间，并且又恰好在两个交点之一的附近。这里的“附近”二字十分关键。为什么是“附近”而不必正好在交点上呢？一个主要的原因是地球是一个庞然大物。这就好比用一只很大的箩筐去兜别人扔过来的东西，东西哪怕扔歪一点，也照样能兜到。日食的形成也一样，由于地球很大，月球即使偏离交点，只要偏得不太过分，影子就仍能投射到地球上，从而仍能形成日食。那么，多大的偏离才不过分呢？计算表明，只要偏离在 15.35° 以内，就能确保日食；如果偏离在 9.92° 之内，则不仅能确保日食，而且还能确保日全食；甚至当偏离在 $15.35^{\circ}\sim 18.52^{\circ}$ 之内时，仍有一定的可能性会发生日食（这时是否发生日食将取决于月球及太阳离地球的远近

等其他因素），这个允许偏离的最大范围所对应的时段被称为食季（eclipse season）。

由于偏离可以发生在交点的任何一侧，因此食季的角度范围是 18.52° 的两倍，即 37° 左右。在地球绕太阳的公转中，走过这样一个角度范围所需的时间约为37天，这就是食季的长度。由于这个长度比一个“月”来得长，因此在一个食季中，至少会发生一次日食。运气好的话，月球有可能会两次经过能形成日食的位置，这样我们就能在一个食季中看到两次日食。比如2011年的6月1日和7月1日这两个同属一个食季的日子就都出现了日偏食（可惜中国不在那两次日偏食的日食带上）。由于在一个食季中有一定的可能性出现两次日食，因此每个世纪的日食次数要比食季本身的数量略多，而且各个世纪的日食次数不尽相同。

现在让我们来介绍一些更精确的日食规律。我们已经知道，日食的发生是因为月球遮挡了太阳。很明显，这时的月相必定是新月（因为太阳在月球的正后方）。因此，日食发生的粗略条件虽然取决于食年和食季，确切日期却是由新月决定的（读者如果想要推算日食的时间，这一点是很重要的）。^[4]在太阳、地球和月球这一系统中，非常巧合的是：19个食年几乎恰好等于223个朔望月——两者都很接近6 585天。这说明与任何一次日食相隔6 585天，就几乎铁定会有另外一次日食。更巧合的是，月球每两次经过近地点（或远地点）的时间间隔——即所谓的近点月（anomalistic month）——为27.554 55天，^[5]239个近点月也几乎恰好就是6 585天。由于近点月描述的是月地距离——从而也是月球的角直径——的变化周期，而月球的角直径是影响日食种类及持续时间的重要因素（因为月球的角直径越大，就越容易形成日全食，而且能遮挡太阳的时间也就越长）。因此6 585天是食年、朔望月和近点月的共同倍数这一巧合，意味着每隔6 585天，与日食有关的太阳、地球和月球的

各种位置参数都几乎恰好彼此相同，由此导致的结果是出现两次几乎一模一样的日食！6 585天这个周期由于这一特点而变得很奇妙，它有一个专门的名称，叫做沙罗周期（Saros），这个名称来自古巴比伦，它的希腊文含义是“重复”。古巴比伦人早在两千五百多年前就知道了这一周期。由于日食的确切日期取决于新月，而新月之间的间隔是朔望月，因此沙罗周期的确切长度由223个朔望月所确定，比较精确的数值是6 585.32天，或18年零11天又8小时。[\[6\]](#)

作为应用沙罗周期的例子，我们可以看一看2009年7月22日经过中国长江流域的，被称为“最壮观日食”的日全食，那次日食的全食时间长达6分39秒，属于超长日全食。与它以沙罗周期相间隔的前三次日食分别发生在1955年6月20日，1973年6月30日和1991年7月11日，也都是日全食，全食时间分别为7分8秒，7分4秒和6分53秒，与当前天文条件下日全食持续时间的理论上限7分31秒（请读者想一想，这一理论上限在什么条件下可以实现）相差不多，也都是超长日全食，彼此间非常相似。不仅如此，我们还可以很放心地预言，与那次日食往未来方向间隔一个沙罗周期的2027年8月2日，也一定会发生一次持续时间很长的日全食（瞧瞧，现在咱们也能预言日食了）。

利用沙罗周期的特点，人们对日食进行了归类，将被沙罗周期所分隔的各次两两相似的日食归为同一序列，称为沙罗序列（Saros cycle）。每个沙罗序列有一个序列号，比如2009年7月22日那次日食所属的序列号是136，即沙罗序列136（Saros cycle 136）。那是一个著名的沙罗序列，它之所以著名，除了当前的全食时间特别长以外，还有一个更重要的原因，我们将在本章的末尾予以揭晓。

有关沙罗序列，还有几个重要特点值得一提。细心的读者也许注意

到了，在沙罗周期中，除了18年零11天这个大头外，还有一个8小时的零头。这个零头虽然不大，却不容忽视。由于地球每24小时自转一周，因此这8小时的零头意味着同一个沙罗序列中，相邻的两次日食在经度上大致相差1/3个地球，即 120° ，比如前面提到的经过中国长江流域的沙罗序列136日食，它的上一次日食发生在中美洲和太平洋上，下一次则将移师非洲北部。同一地点的观察者是无法观察到沙罗序列中的相邻日食的（从这个角度上讲，古巴比伦人能够发现沙罗周期实在是一件很奇怪的事情，感兴趣的读者请想一想，或查查资料，探究一下他们是如何发现沙罗周期的）。有关沙罗周期的另一个重要特点是它具有一定的寿命，而不能千秋万载。这是因为虽然19个食年的长度非常接近223个朔望月，但毕竟不是完全相等。仔细的计算表明，前者比后者长了0.46天。由此导致的后果，是经过一个沙罗周期后，月球相对于升交点或降交点的位置将会偏离 0.48° 。我们在前面说过，日食食季的角度范围约为 37° ，因此经过不到80个沙罗周期，月球位置的总偏移就会超过食季的角度范围，这意味着沙罗序列的终结，这个过程的时间约为1 300多年。因此沙罗周期是不能千秋万载的——确切地说是只能千秋，不能万载。

日食的发生还有其他一些规律，限于篇幅，我们就不介绍了。对日食规律的研究不仅能让天文学家们预言未来日食出现的时间、地点及持续时间，而且还能帮助历史学家们还原某些曾经提及过日食的历史事件的准确时间（比如希罗多德所记载的那次因日食而终止的战争就被历史学家们确定为最有可能发生在公元前585年），并且校正古代的历法。对日食规律的推算——如我们在上面所做的——从原理上讲虽然是直截了当的，但要想做得精确却很不容易，需要考虑各种细微引力摄动的影响。以对日食带的推算为例，由于月球阴影在地球表面大部分区域的运动速度高达每秒几百米，几分钟的时间误差就足以造成上百千米的空间

误差。依据这样的预言去追逐日食，就会像革命大片中那些智力低下的敌军一样，完全扑错地方。

在接下来的篇幅中，我们简单介绍一下日食的种类。如图4.2所示，在日食中，如果月球的本影扫过地球，相应的日食就称为日全食，被本影扫过的区域称为全食带；如果本影无法触及地球，但其延长区——即所谓的束后本影（*antumbra*）——扫过地球，相应的日食就称为日环食（*annular eclipse*），被束后本影扫过的区域称为环食带；如果本影和束后本影都无法触及地球，而只有半影扫过地球，相应的日食就称为日偏食（*partial eclipse*），被半影扫过的区域称为偏食带。显然，在日全食发生时，月球的角直径必须大于太阳，而在日环食发生时，月球的角直径必须小于太阳。这两种情形都能发生，是因为我们在第2章中提到过的一个事实：即“在太阳和月球之间存在一个美妙的巧合，那就是它们看起来几乎是一样大的”。这一巧合，加上月球和地球的公转轨道都是椭圆，从而月地和日地距离都会随时间改变这一特点，使得月球的角直径有时候比太阳大，有时候比太阳小。这是我们有时候能看到日全食，有时候能看到日环食的根本原因。

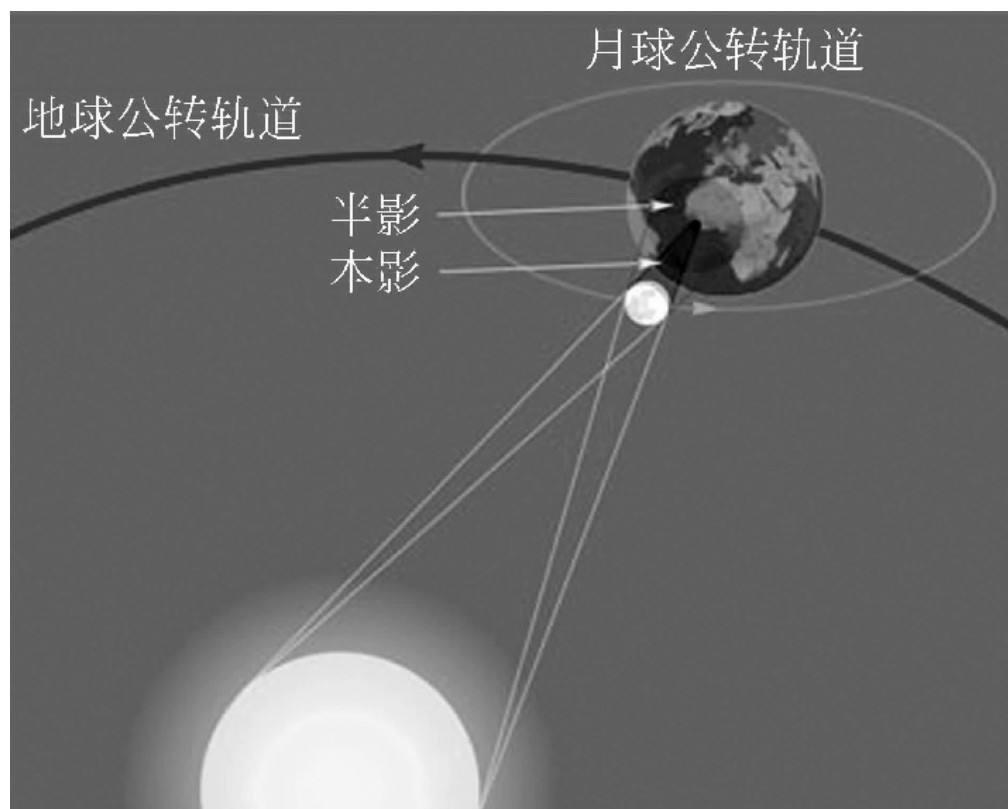


图4.2 日食的图示

由于无论月球的角直径比太阳大还是小，两者终究都很接近，而地球表面又是弧形的，因此偶尔会出现一种更罕见的日食，那就是地球表面的某个区域离月球较近，使得月球看上去比太阳略大，从而形成日全食。而该区域之外的某个其他区域离月球较远，使得月球看上去比太阳略小，从而形成日环食。这种同时具有全食和环食的特殊日食被称为全环食（hybrid eclipse）。在20世纪的224次日食中，日全食、日环食、日偏食和全环食的次数将分别为68、72、77和7次。如此种类繁多的日食出现在同一颗行星上，这不仅在太阳系中是绝无仅有的，甚至在任何其他恒星的周围，其形成概率也是微乎其微的。从这个意义上讲，我们在地球上所能看到的日食是一种极为稀有的天象奇观。事实上，即便对地球来说，这种奇观也不是任何时候都存在的。日食带的狭窄（全食和环食带的最大宽度都只有两三百千米）说明月球与地球的距离只要再大一

点，日全食就将不复存在，只要再小一点，日环食就会烟消云散。而天文学家们早就知道，月球与地球的距离一直都在增加（目前的增加速度为每年38毫米），因此早年的地球上不存在日环食，未来的地球上则不会有日全食。从地质学时标上讲，目前这么丰富多彩的日食格局即使在地球上也是我们这个时代独有的。

在结束本章前，我们还要补充一点，算是对本章标题的注释。我们在本章中虽然只讨论了日食，但前面提到某些概念，比如食季，是同时适用于日食和月食的。食季既是日食的摇篮，也是月食的温床，在一个食季中，月球挡在太阳和地球中间就是日食，地球挡在太阳和月球中间则是月食。那么这两种情形哪一种更容易发生呢？仔细的分析表明，是日食，日食的食季长度要明显大于月食。因此，在地球上所能看到的日食次数要明显多于月食。^[7]从这个意义上讲，日食在大范围上虽然很稀有，在如今我们这个小小星球上却很寻常，因此是一种“既寻常又稀有的奇观”。

读到这里，有些读者也许会感到困惑，因为在多数人的印象里，日食远比月食稀罕。之所以会有这种印象，是因为月食发生时，只要天气许可，半个地球上的人都可以看到，而日食发生时，即便老天开眼，也只有一个狭长日食带上的人才会有眼福，而且这日食带还常常落入海洋、沙漠、雪山、极地等人烟稀少的地区。因此虽然从总次数上讲日食比月食更多，但除非你不惜跋山涉水去追逐日食，否则你所能看到的日食会明显少于月食。尤其是，假如你特别眷恋家乡，一步都不愿离开的话，也许一辈子都未必能看到一次日食（但依然可以看到几十次月食）。以日全食为例，在地球表面随机选定的地点上，平均每隔410年才能看到一次。^[8]

日食——尤其是日全食——不仅是一种美丽的天象奇观，在天文学家们的眼里，更是研究太阳的绝佳机会。对于天文学上的多数研究来说，研究对象的过于暗淡是天文学家们面对的主要困难之一，但对于太阳研究来说，情况恰恰相反，过于明亮才是最大的“麻烦制造者”。太阳的明亮甚至使得天文学家们不敢用精密的哈勃望远镜观察离太阳最近的水星。在这种情况下，日食的存在对于太阳研究有着极大的重要性。这种重要性早在两百多年前就得到了令人瞩目的确立。在1780年10月27日的日全食期间，美国独立战争的交战双方（英、美）为了让哈佛大学的天文学家们能在全食带上进行科学研究，破天荒地作出了停战一天的决定（可惜天文学家们不争气，竟然算错了日食带，以至于功败垂成）。

最后，我们要兑现诺言，来揭晓前面提到过的使2009年7月22日经过中国长江流域的日全食所属的沙罗序列136著名的“更重要的原因”。这个原因很简单：如果我们将2009年7月22日往前推5个沙罗周期，即32927天，我们将得到一个日子1919年5月29日。在那个日子里也发生了一次属于沙罗序列136的日全食，它的持续时间也很长：6分51秒。熟悉物理学史的读者看到这个日期也许已经想起来了，是的，在那一天英国天文学家爱丁顿（Arthur Eddington, 1882—1944）对广义相对论中光线的引力偏折效应进行了检验。那次检验也许是日全食期间所有科学研究中最著名的一次，正是它使得沙罗序列136更加出名。

[\[1\]](#)在阿西莫夫所设想的这种多星系统中，由于存在极为复杂而且巨大的引力摄动，行星实际上几乎不可能维持周期或准周期轨道，从而几乎不可能为生物进化提供足够稳定的环境。在人们探讨什么样的行星上有可能演化出高等生物时，双星、三星或多星系统中的行星基本上是可以被直接排除在外的。因此阿西莫夫在那篇小说中设想的环境只是一种虚构。

[2]这指的仅仅是普通情形，在某些特殊情形下，月球的阴影不是自西向东掠过地面，而是从空中直接压向地面的。感兴趣的读者可以想一想，在什么情况下月球的阴影会从空中压向地面？另外也请想一想，月球的阴影自西向东掠过地面，与月球的东升西落是否矛盾？

[3]如果地球的直径不是12 700千米，而是90000千米以上，月球的影子（包括本影和半影）就总能投射到地球上。那样的话，日食就真的会大约每个月都出现一次。

[4]作为用新月推算日食确切日期的例子，让我们推算一下2010年1月15日经过中国四川、湖北等地的日环食之后的下一次日食的时间。由于两个食季之间的平均间隔（即食年的一半）是173.3天，很接近6个朔望月（约177天），考虑到食季的长度有37天，而日环食和日全食通常位于食季的中部，因此我们可以预期，2010年1月15日的日环食之后相隔6个朔望月的那一天，即2010年7月11日，会有一次日食。事实上也的确如此，那一天有一次日全食，只可惜它的日食带主要在南太平洋上，只有一小段经过南美洲南端。感兴趣的读者请将2010年1月15日这个日子往回推6个朔望月，您将得到我们在本节中介绍过的另一次日食。

[5]“近点月”与“恒星月”的差别来自于月球近地点（或远地点）的进动，这种进动与月球公转方向相同，周期为8.85年（读者可以根据此定量验证一下两者的关系）。

[6]想要验证沙罗周期的读者可以采用以下数据：1食年=346.620 077天，1朔望月= 29.530 59天（近点月的数据在正文中已经给出了）。这里要顺便提醒读者的是，如果一个沙罗周期中的18年间恰好有5个而不是4个闰年，则沙罗周期的历法表示将是18年零10天又8小时。

为了行文简洁，后文将不再提醒这一点。另外还有一点要提一下，那就是日食的种类及持续时间与地球到太阳的距离也有关系（不过这个关系比较次要，因为地球公转轨道的椭率远小于月球）。沙罗周期只比18年多10或11天（即0.03年）这一特点，意味着相隔一个沙罗周期后，地球与太阳的相对位置也几乎不变。

[7] 早在1887年，奥地利天文学家奥普泽（Theodor von Oppolzer, 1841—1886）就对从公元前1208年到公元2161年这3 369年间的所有日食和月食进行了计算，结果是日食约有8 000次，月食只有约5 200次。

[8] 更具体地说，北半球的日全食多于南半球。在一个随机选定的北半球地点上平均每隔330年就会有一次日全食，而南半球的这一间隔长达540年。产生这种差异的原因，是因为地球在北半球的夏季里离太阳较远（位于远日点附近），这时太阳的表观直径较小，容易形成日全食。而恰恰是在这段时间里，北半球被太阳照到的范围大于南半球（因为是北半球的夏季），从而在日全食最容易形成的时段里有机会看到更多的日全食。不过北半球并不能永远保有这一优势，由于地球自转轴和地球公转轨道的进动，这一优势将在几千年之后逆转。



绘画：张京

5 插曲：爱丁顿在1919

我们在第4章末尾提到，英国天文学家爱丁顿利用1919年5月29日的日全食，对广义相对论中光线的引力偏折效应进行了检验。在本章中，我们就来聊聊这段或许是日全食期间所有科学研究中最著名的往事，作为太阳故事的插曲。

这段往事既然与验证广义相对论有关，当然得从广义相对论的“始作俑者”爱因斯坦（Albert Einstein，1879—1955）说起。1907年11月，爱因斯坦正在赶写一篇有关狭义相对论的综述。^[1]他一边写，一边思考着一个棘手的问题，那就是如何将狭义相对论推广到非惯性系中。结果有一天，他产生了一个灵感。他在后来访问日本期间所做的一次演讲中回忆了当时的情形，他说当时他正坐在伯尔尼专利局的办公室里，忽然一个念头闪了出来，“如果一个人自由下落，他将不会感觉到自己的体重”。这个念头就是著名的等效原理（equivalence principle）的雏形。按照等效原理，引力场和加速场在小范围内是不可分辨的，由此导致的一个必然结果，是光线会在引力作用下偏折（因为做加速运动的观测者可以让光线看起来是弯曲的）。1911年，爱因斯坦对这种光线的引力偏折效应进行了定量计算，结果发现掠过太阳表面的光线会偏折 $0.83''$ （arc second）。^[2]

爱因斯坦的这一计算虽然在很多环节上都打着相对论的旗号，其实与相对论并无必然联系，把光当成在牛顿引力场中运动的普通物质也能得到同样的结果。事实上，后来有人从对历史文献的“考古”中发现，早在1808年，德国科学家索德纳（Georg von Soldner，1776—1833）就得

到过同样的结果。甚至比那更早，1784年，著名的英国科学家卡文迪许（Henry Cavendish, 1731—1810）就在一篇未发表的手稿中指出过星光在牛顿引力场中会弯曲。当然，爱因斯坦并不知道那些陈年旧账，对他来说，那是等效原理的推论，是一个新的结果。

显然，如果光线在引力场中会偏折，那么当一颗星星出现在太阳近旁时，它的实际方位与表观方位之间就会出现偏差（图5.1）。这个偏差原则上是可以通过比较太阳不在视线附近时星星的方位与太阳在视线附近时星星的表观方位之间的差异来进行测量的。但这里有一个问题，那就是太阳实在太抢眼了，一出场就会使得“六宫粉黛无颜色”——整个天空基本上就看不到星星了，更别说是它的“卧榻之侧”了。怎么办呢？只有等待机会。等待什么机会呢？等待日全食的机会。因此，爱因斯坦在1911年论文的末尾写道：“由于日全食期间位于太阳附近那部分天空中的星星会变成可见的，理论的这一推论是有可能被观测到的。”他并且呼吁：“非常希望天文学家们能过问此处所提的问题，即使上述考察看起来显得根据不足或冒险行事。”

最早被爱因斯坦的呼吁打动并愿意付诸行动的，是在柏林天文台“打杂”的一位名叫弗洛因德利希（Erwin Freundlich, 1885—1964）的德国天文学家。这位曾跟随著名数学家克莱因（Felix Klein, 1849—1925）学习数学的年轻人当时正干着一些绘制星表之类“嘴里都快淡出鸟来了”的乏味工作。1911年8月的某一天，事情有了一点转机，他接待了一位来自布拉格（Prague）的物理学家。这种接待工作原本倒也不是什么美差，因为轮得到他去接待的人一般是不重要的。但此次接待的物理学家稍有些不同，他本身虽然不重要，却恰好认得一位有点重要——并且正在变得越来越重要——的人物：爱因斯坦。闲聊中，那位物理学家向弗洛因德利希传达了爱因斯坦1911年的论文精神。弗洛因德利希立

刻就对爱因斯坦的预言产生了兴趣，因为那比他当时正在做的任何事情都更有意思。于是他与爱因斯坦建立了通信联系，讨论如何验证光线的引力偏折。

当然，验证的方法爱因斯坦已经说了，是利用日全食。但问题是，日全食的机会不是经常有的，即便有也往往得跋山涉水前往日食带才能观测，这就涉及了钱的问题。那么，有没有不花钱的办法呢？爱因斯坦出了一个点子，他建议弗洛因德利希找一些以前日全食期间拍摄的旧相片，看能否有所发现。弗洛因德利希采纳了这个点子，向世界各地的天文学家索要了一些日全食期间拍摄的旧相片。可惜结果很令人失望，那些相片无一具备验证光线偏折所需的清晰度。

没办法，只能花钱了。

弗洛因德利希把目光瞄准在了1914年8月21日的日全食上。那是一次属于沙罗序列124的日全食，全食时间约为2分14秒，全食带从西亚延伸到北欧，其中离德国较近的观测点在俄国。弗洛因德利希决定多管齐下：一方面与其他天文学家联系，希望届时能分享他们的相片；另一方面自己也筹集了经费，前往俄国亲自进行观测。事情的筹划虽不无困难，但总算是办下来了。但他没有料到的是，随着第一次世界大战的爆发和蔓延，1914年8月1日，德国对俄国宣战了。这一来，已经抵达俄国的弗洛因德利希等人的身份立刻发生了戏剧性的变化，由国际友人变成了敌国奸细。俄国人民的眼睛是雪亮的，很快就把他们这一小撮带着可

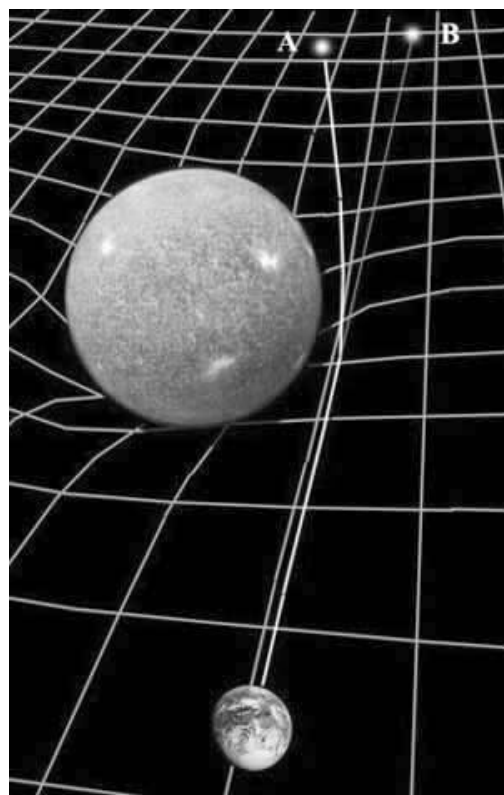


图5.1 太阳近旁的星光偏折

疑仪器的德国人扣留了下来。所幸的是，他们后来被作为战俘交换回了德国，但试图验证广义相对论的光线引力偏折效应的努力就此付诸了东流，^[3]这是爱丁顿故事之前的主要序曲。

这段序曲中比较微妙的是，在1914年的时候，爱因斯坦虽然正在他的大学同学格罗斯曼（Marcel Grossmann, 1878—1936）的帮助下快速推进着广义相对论的研究，他对太阳近旁光线引力偏折角度的预言却仍停留在0.83角秒上，只有正确值的一半。一些注意到这一历史细节的传记或科普作品喜欢就这一点进行发挥，认为假如弗洛因德利希能如愿以偿地拍摄1914年的日全食，爱因斯坦的理论很可能会遭到一次观测的否决。其实就当时的情形来说，这个可能性是很小的。因为一方面，当时的观测误差很大，要在0.83角秒和两倍于此的正确值之间作出无争议的区分是不容易的。另一方面，人们从弗洛因德利希先前对日全食旧相片的处理中发现，他的分析手段存在缺陷，与后来人们普遍使用的手段相比，具有更大的误差。除这两点外，还有一个因素也不容忽视，那就是在光线偏折的正确值尚未出炉的时候，人们并未对如今称为“牛顿值”的这个错误结果与广义相对论的结果进行区分，他们更关心的是光线偏折到底是有还是无。因此只要观测证实偏折存在，哪怕具体数值有一定出入，考虑到误差很大，也依然会被视为对爱因斯坦理论的重大支持。因此，那次观测的流产并不像一些作者所渲染的那样富有戏剧性。

现在回到爱因斯坦这边来。1915年底，爱因斯坦完成了广义相对论，并将太阳近旁光线的引力偏折角度修正为1.7角秒，即原先结果的两倍。此时的爱因斯坦虽远没有后来那样的公众知名度，在学术界却已是重量级人物了。1913年，柏林大学为了把他从苏黎世挖到柏林，不惜派出了像普朗克（Max Planck, 1858—1947）和能斯特（Walther Nernst, 1864—1941）那样史上最牛的猎头阵容，动用了包括普鲁士科

学院院士在内的超级诱饵，才告得手。

虽然战争还在继续，爱因斯坦所在的柏林几乎成了与世隔绝的孤城，但爱因斯坦在广义相对论上所取得的重大进展，还是很快就通过中立国荷兰的物理学家洛伦兹（Hendrik Lorentz, 1853—1928）和艾伦菲斯特（Paul Ehrenfest, 1880—1933）传到了一些感兴趣的人那里，其中包括莱顿天文台（Leiden Observatory）的台长德西特（Willem de Sitter, 1872—1934）。德西特一直对爱因斯坦的工作怀有浓厚兴趣，是除爱因斯坦本人以外最早研究广义相对论的人之一。为了让英国同行们也能分享爱因斯坦的工作，他将一份爱因斯坦的论文寄给了英国皇家天文学会。而当时任皇家天文学会秘书的不是别人，正是爱丁顿。德西特寄来的论文第一时间就落到了他的手里。

就像把大陆和台湾隔开的不仅仅是台湾海峡，把英国科学界和德国科学界隔开的也并不仅仅是英吉利海峡。两百多年来，这两个国家的科学界之间一直存在着嫌隙（牛顿-莱布尼茨之争留下的恶果），而且当时英国和德国正处于交战状态，情况更是雪上加霜。但爱丁顿却是一个例外，他是一位坚定的和平主义者，丝毫不受英、德两国之间任何争斗的影响，对德国科学家及其工作没有任何成见。爱因斯坦在一篇阐述广义相对论场方程的通信中曾经表示：“任何人只要对这一理论有着充分理解，就很少能从它那不可思议的魔法中逃脱出来。”巧得很，爱丁顿恰好就具有使自己“对这一理论有着充分理解”所必需的数学功底，他的命运也就可想而知了。在此后两年多的时间里，爱丁顿被广义相对论那“不可思议的魔法”所彻底吸引，成为了当时为数不多通晓并积极传播广义相对论的物理学家之一。^[4]而广义相对论对光线引力偏折的预言自然也引起了他的极大兴趣。



英国天文学家爱丁顿（1882—1944）

1918年，爱丁顿开始与皇家天文学家、格林尼治天文台（Royal Greenwich Observatory）台长戴森（Frank Dyson, 1868—1939）商量组织日食远征队，检验广义相对论的光线引力偏折效应的计划。这个计划当时对爱丁顿来说有着双重意义，一方面当然是可以检验让他如此着迷的广义相对论；另一方面，则可以使他摆脱一个现实困境。我们在前面提到，爱丁顿是一位和平主义者，事实上，他的和平主义立场还相当激进，激进到了因拒服兵役而将自己推到牢狱之灾边缘的程度。在这个节骨眼上，检验广义相对论成为了戴森帮他在英国海军部面前开脱的最好理由。^[5]

方略既已确定，接下来就是寻找合适的日全食了。我们在第4章中曾经说过，对单一地点来说，日全食的出现是相当稀有的。拿英国来说，当爱丁顿想要检验广义相对论时，英国本土已有近两百年没有发生日全食了（英国本土的上一次日全食发生在1724年），而且未来的短时间内也不会有。爱丁顿如果想在英国本土进行观测，起码得再等上9年，因为英国本土只有到1927年才会有新的日全食，而且持续时间只有24秒，实在是“匆匆，太匆匆”。因此爱丁顿选了一个更近并且更好的时机，这个时机就是我们在第4章末尾提到的发生于1919年5月29日，属于沙罗序列136，持续时间长达6分51秒的超长日全食。那样的日全食对观测来说无疑是极为有利的。而更有利的则是，那时的太阳将位于包含许多亮星的金牛座（Taurus）毕星团（Hyades）。这样的机会哪怕在全世界范围内也不是常有的。那次日全食唯一不利的条件，是全食带远在赤道附近，横贯非洲、大西洋及南美洲，却远离欧洲（图5.2）。在对几个候选地点进行气候分析之后，爱丁顿在非洲和南美洲各选了一个观测点，其中非洲的观测点选在非洲西海岸附近的小岛普林西比（Principe），南美洲的观测点则选在了巴西亚马逊丛林中的小镇索布拉（Sobral）。



图5.2 1919年日食的全食带

不过，由于这两个地点远离英国本土，爱丁顿的计划能否实现，还得看战争能否及时结束。只有战争结束了，远征队才能平安穿越大西洋（否则会受到德国潜艇的威胁）。幸运的是，第一次世界大战的战火在燃烧了四年，付出了1 500万条人命的代价后，终于在1918年11月熄灭了。1919年3月，两支英国远征队如期出发，其中爱丁顿亲自率领的一支前往非洲的普林西比，另一支前往南美洲的索布拉。观测的过程想必大家早已在其他科普作品中读到过，本文就不赘述了，基本上就是环境是如何如何的闷热不堪，蚊虫是如何如何的“毁人不倦”，天气是如何如何的忽好忽坏，心情是如何如何的“此起彼伏”（宋丹丹语）。最终的结果是两支考察队都拍到了相片，满载而归。

接下来就是数据分析。爱丁顿后来很喜欢讲述的一个故事，是说远征队出发前他的助手曾经问戴森，假如观测到的结果是爱因斯坦预言值的两倍，会怎么样？戴森回答说，如果那样的话，爱丁顿就会发疯，而你只好一个人回来了。其实，主要的数据分析是在回英国之后才做的，爱丁顿即使要发疯也得回英国来发。由于数据比较微妙，分析过程持续

了较长时间。具体地说，索布拉远征队使用了两类不同的观测方法，一类用的是一架4英寸折射望远镜，拍摄的相片共有8张（其中一张因云层干扰而没能拍到星星，因此有效相片为7张），这些相片显示的偏折角度为 $1.98''$ ，与爱因斯坦的预言比较接近，误差也相对较小。另一类用的是天体照相机（astrograph），拍摄的相片共有19张，这些相片显示的偏折角度约为 $0.9''$ ，与“牛顿值”比较接近，但这些相片的图像比较模糊，据怀疑是因为仪器中的定天镜（coelostat mirror）受热变形所致。爱丁顿自己那组用的也是天体照相机，共拍摄了16张相片，但其中只有两张质量较好，经分析得到的偏转角度为 $1.61''$ ，也比较接近爱因斯坦的预言。

那么结论是什么呢？爱丁顿作出了自己的选择，他以仪器有问题为由丢弃了 $0.9''$ 那组比较接近“牛顿值”的结果，而保留了索布拉远征队的 $1.98''$ 的结果，以及他自己那组其实也不怎么靠得住的 $1.61''$ 的结果。这两组结果的平均非常接近爱因斯坦的预言。因此爱丁顿的结论是广义相对论对光线偏折的预言得到了证实。

1919年11月6日，英国皇家学会和皇家天文学会举行联合会议，正式宣布了爱丁顿的观测结果及结论。会议由电子的发现者，皇家学会主席，著名实验物理学家汤姆孙（J. J. Thomson, 1856—1940）主持。在巨幅的牛顿画像前，戴森报告了观测结果，他表示，在仔细研究了相片之后，他认为它们毫无疑问地证实了爱因斯坦的预言。汤姆孙主席基本接受了这个乐观结论，表示这是自牛顿以来有关引力理论最重要的结果，是人类思维的最高成就之一。^[6]特意从剑桥赶来参加会议的著名哲学家怀特黑德（Alfred Whitehead, 1861—1947）后来在自己的著作《科学与当代世界》（*Science and the Modern World*）中回忆了当时的情形，他写道：

那种洋溢着浓厚兴趣的气氛完全是希腊戏剧式的。我们都齐声称颂着这一卓越事件在发展过程中所显示出的命运的律令。舞台本身就充满了戏剧性：传统的仪式和后面悬挂着的牛顿画像都在提醒我们，最伟大的科学成就在两个多世纪后的今天第一次得到了修正。

这无疑是科学史上的一个著名时刻，不过在那之前，爱丁顿的观测结果就在一个小范围内传开了。早在9月22日，洛伦兹就已经发电报将消息告诉给了爱因斯坦。稍后，10月4日，普朗克向爱因斯坦表示了祝贺。10月22日，普鲁士科学院院士，德国哲学及心理学家斯顿夫（Carl Stumpf, 1848-1936）也向爱因斯坦表示了“最诚挚的祝贺”，并表示，在经历了军事和政治的失败后，德国科学能够取得这样的胜利令人感到自豪。而爱因斯坦在接到洛伦兹的电报后随即将消息转告给了已罹患重病的母亲（他母亲在几个月后就病逝了）。[\[7\]](#)

很多科普或传记作品在描述爱因斯坦得知爱丁顿观测结果后的反应时，喜欢渲染他的自信与超脱。毫无疑问，爱因斯坦对广义相对论有着极强的信心，但再好的理论也必须经过观测的检验。因此任何物理学家只要还在关心作为研究对象的大自然，就无法在一个重大的观测检验面前保持超脱，爱因斯坦也不例外。事实上，爱因斯坦一直非常在乎天文学家们对广义相对论的验证。拿光线的引力偏折来说，他不仅反复呼吁天文学家们对此进行检验，而且还在弗洛因德利希为1914年的远征筹集经费时，一方面动用自己的影响力给予帮助，另一方面表示若有必要，将亲自出资支持弗洛因德利希的观测。而在爱丁顿的结果出炉前，爱因斯坦也不止一次在给亲友的信中询问观测结果，或流露出等待结果的急切。

爱丁顿在1919年的故事大致就是这些。他后来把获得观测结果的那一刻称为自己一生最伟大的时刻。那一刻不仅是他个人的伟大时刻，而且也使爱因斯坦几乎在一夜之间获得了世界性的公众影响。但这些耀眼的光环并未让学术界失去应有的沉稳，对于爱丁顿处理观测数据的方法，很快就有人提出了质疑。由于篇幅关系，本文就不展开讨论那些质疑了。不过要指出的是，爱丁顿的论文如实列出了所有的观测数据，其中包括被他以仪器有问题为由丢弃的数据。他是在公布了全部数据的基础上进行自己的分析与取舍的。因此人们对他的数据取舍的合理性虽然可以有各种看法，但如果把那种取舍渲染成类似于舞弊的行为则是很不恰当的。另外要指出的一点是，1978年，当爱因斯坦诞辰一百周年的日子即将到来时，当时任格林尼治天文台台长的默里（Andrew Murray）与同事哈维（Geoffrey Harvey）等人用包括计算机处理在内的现代手段重新分析了被爱丁顿丢弃的那些相片，结果发现它们所记录的星光偏折其实不是当年以为的 $0.9''$ ，而是 $1.55 \pm 0.34''$ 。因此即使是那些相片，也与爱因斯坦的结果更为接近。

当然，更重要的是，无论是广义相对论，还是其他科学理论，也无论已经出现了多么具有轰动效应的检验，新的检验都永远不会停息。在爱丁顿之后，更多的人对广义相对论产生了兴趣，在后来的日全食期间，更多的天文学家加入了检验的行列。再往后，随着技术的发展，日全食已不再是检验星光偏折的唯一机会，检验的精度也由原先的百分之几十提高到了万分之一量级。直到今天，人们依然在用各种手段检验着广义相对论的各种预言，科学就是这样一种不断自省、永不停息的努力。

[1] 他的这篇题为《关于相对性原理和由此得出的结论》的综述是应德国《放射性与电子学年鉴》的编辑斯塔克（Johannes Stark, 1874

—1957) 的约稿而写的。斯塔克是德国实验物理学家，斯塔克效应 (Stark effect) 的发现者，1919年的诺贝尔物理学奖得主。他后来成为“德意志物理学”的狂热鼓吹者，对爱因斯坦展开了激烈攻击。但在1907年，两人的通信是完全彬彬有礼的，斯塔克对当时还在专利局的爱因斯坦的称呼是“非常尊敬的同事”，爱因斯坦对斯塔克的称呼则是“非常尊敬的教授先生”。

[2] 这是爱因斯坦1911年原始论文所给出的数值，利用现代天文数据，爱因斯坦当年结果所对应的数值应为 $0.87''$ ，完整的广义相对论预言则是 $1.75''$ 。下文在引述各种当年所采用的数值时，将不再一一注释。

[3] 不过即便没有战争的干扰，他们的观测也不可能成功，因为那天的全食期间太阳恰好被厚厚的灰云所遮盖。打算与弗洛因德利希合作的美国里克天文台 (Lick Observatory) 台长坎贝尔 (William Campbell, 1862—1938) 带领的美国远征队当时也在那里，他们虽然未被扣留，却也因为天气原因而无功折返 (返回时因战争原因，将仪器暂时留在了俄国)。坎贝尔后来在1918年6月8日的属于沙罗序列126的日全食期间再次进行了观测 (那次的全食带经过美国)。但可惜他留在俄国的仪器因战争原因尚未运回，只能用一些临时拼凑的仪器，而当时的天气状况也不佳，拍到的相片质量很差。此外，在弗洛因德利希之前，1912年10月10日的属于沙罗序列142的日全食期间，曾有一支阿根廷远征队前往巴西进行观测，结果因遭遇雨天而失败。

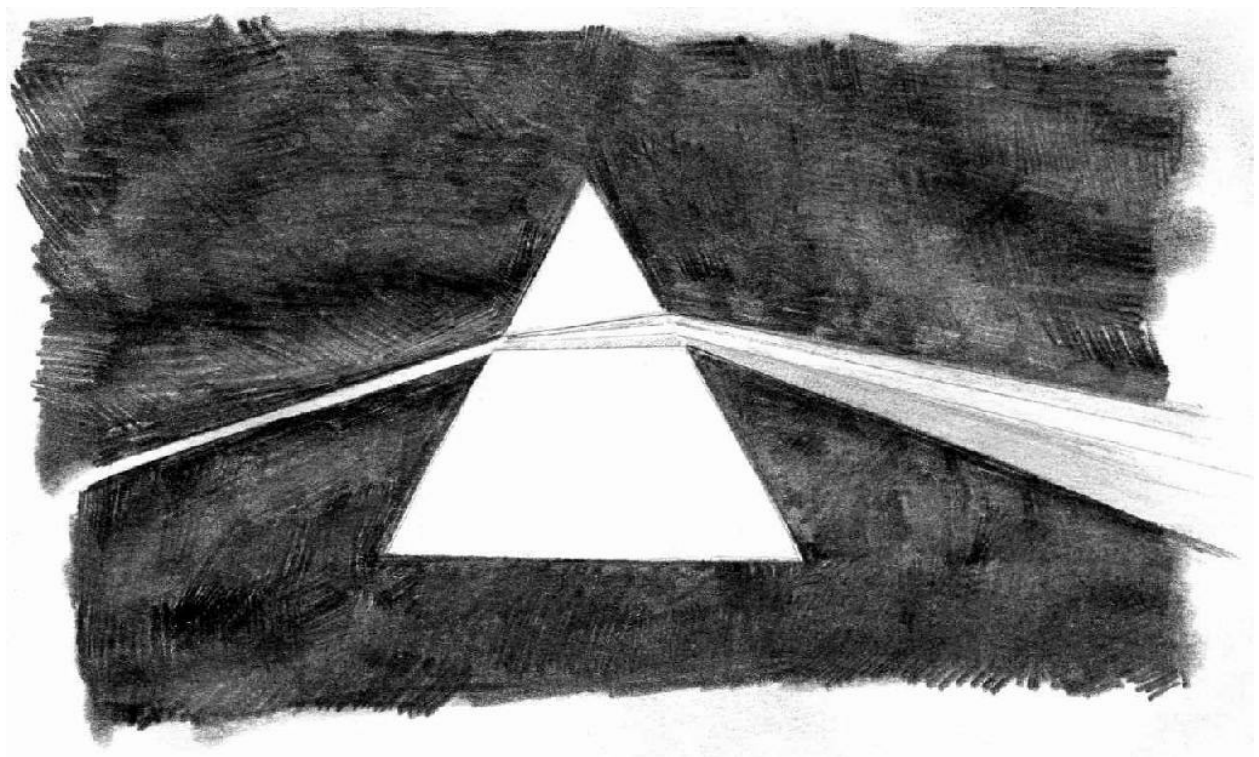
[4] 有一个流传很广的故事说波兰人西尔维斯坦 (Ludwik Silberstein, 1872—1948) 曾对爱丁顿说世界上只有三个人懂得广义相对论，而爱丁顿是其中之一。爱丁顿不置可否，当西尔维斯坦让他不要太谦虚时，他说：“不，我只是在想那第三个人是谁。”这个故事印

度裔美国物理学家钱德拉塞卡（Subrahmanyan Chandrasekhar, 1910—1995）曾听爱丁顿亲口说过，因此应该是真的，只不过这个故事对广义相对论的难度显然有所夸大。事实上，即便在那时，除爱因斯坦和爱丁顿外，懂广义相对论的起码还应该算上希尔伯特（David Hilbert, 1862—1943）、外尔（Hermann Weyl, 1885—1955）、德西特等人。另外一个也流传很广的传闻是美国《纽约时报》散布的，说爱因斯坦将自己的某部文稿交给出版商时警告对方说世界上理解他的人不超过十二个。这个故事的真实性不得而知。钱德拉塞卡在他撰写的爱丁顿传记中认为，如果不是当年人们夸大了广义相对论的难度，很多20世纪六七十年代的工作在二三十年代就应该能很容易地实现。对于这个看法，当然是见仁见智，不过广义相对论的难度曾受到很大夸张是不争的事实。

[5]爱丁顿的拒服兵役能被英国当局容忍，还有一个原因是因为不久前的1915年，优秀的英国年轻物理学家莫塞莱（Henry Moseley, 1887—1915）在战斗中阵亡。英国科学界在震惊之余，许多著名科学家出面为爱丁顿请愿，表示让像爱丁顿那样的杰出科学家参军不符合英国的长远利益。

[6]有很多作者在引述汤姆孙的话时，将之改为了“这是人类思想史上最伟大的成就之一——也许是最伟大的成就”，这后半句是加油添醋。当日的会议记录显示，汤姆孙并未说过那样的话。

[7]洛伦兹电报所传达的含义其实相当模糊，它的内容是这样的：“爱丁顿发现了太阳边缘的星光偏折，初步数值在9/10[角]秒及其两倍之间。”从这个电文中很难判断爱丁顿所证实的究竟是爱因斯坦还是牛顿。而爱因斯坦给他母亲的信中也只是笼统地写道：“今天有个好消息，洛伦兹来电报告告诉我英国远征队证实了太阳使光线偏折。”



绘画：张京

6 阳光里的奥秘

在前面几章中，我们介绍了太阳的大小和远近，地心说vs日心说，日食的规律等。那些介绍所涉及的分析手段大都是几何手段。从本章开始，我们将介绍一类全新的手段：物理手段。我们将会看到，那是一类远比几何手段更有效的手段，在它的帮助下，科学家们很快就将太阳研究推向了纵深。这种从几何手段向物理手段的过渡，是与天文学自身的发展脉络基本一致的。因为在历史上，人们对天文学的研究首先是从研究天体运动的几何规律入手的，[\[1\]](#)用物理手段研究天文的所谓“天体物理”（astrophysics）这一重要分支则是在晚得多的时候才出现的（而且它的出现与本章所要介绍的内容有着密切关系）。

天文学沿这样的历史轨迹发展不是偶然的，拿太阳来说，想对它进行深入研究，首先必须克服一个显而易见的困难，那就是它离我们实在太远了（1.5亿千米），而且也实在太热了（表面温度就足以使任何已知的物质气化），我们几乎永远也不可能像研究地球一样到太阳上去钻个孔、挖个洞、采集一些样本。这一现实的困境曾使一些人深感悲观。1835年，实证主义哲学的创始人，法国哲学家孔德（Auguste Comte, 1798—1857）曾经预言，人类永远也不可能了解太阳和星星的化学组成。

幸运的是，与很多其他哲学家曾经发表过的有关科学的高论一样，孔德的预言很快就破灭了。因为太阳虽然很遥远，而且很热，但它却很慷慨地把一样东西送到了地球上，那就是阳光。这个初看起来很寻常的事实有着极不寻常的推论。有了它，我们这个星球才有生命。但阳光带

给我们的不仅仅是生命，还有信息，比如有关太阳化学组成的信息，这是孔德所不知道的。事实上，比孔德的预言早了20年，1814年，德国物理学家夫琅禾费（Joseph von Fraunhofer, 1787—1826）就发明了一种新的仪器，叫做光谱仪（spectroscope），为人们解读阳光里的信息提供了工具。

不过在介绍夫琅禾费之前，我们先要“论资排辈”一下，向大家引见两位前辈。

这两位前辈中的第一位对所有人大概都是“久闻大名，如雷贯耳”的，他就是牛顿（Isaac Newton, 1643—1727）。在17世纪60年代中期，牛顿做过很多光学实验，在其中一组实验中，他让阳光从一个小孔射入屋内，然后经过一个三棱镜，最后投射到一块屏上（彩图5）。他惊讶地发现，出现在屏上的居然是一个色彩缤纷的长椭圆形影像。虽然在牛顿那个时代，人们对光的本性还一无所知，但牛顿毕竟是牛顿，他敏锐地意识到——并且通过进一步的实验证实了——这一现象所揭示的有关阳光的重要性质：那就是阳光是由不同颜色的单色光组合而成的。出现在屏上的彩色影像，则是由于不同颜色的单色光在三棱镜中的偏折角度不同，而被投射到屏上的不同位置所产生的。如今我们知道，牛顿发现的这一现象就是所谓光的色散（dispersion），即不同波长（从而颜色各异）的光在色散介质中的折射率不同，从而偏折角度不同。而他在屏上看到的彩色影像则是最粗糙的阳光光谱——确切地说是阳光光谱中的可见光部分（这部分约占阳光总能量的40%）。牛顿的这一重大发现可以算是一种很原始的光谱分析（spectrum analysis），它是人类在探索光的本性道路上迈出的重要一步。[\[2\]](#)

但是从了解太阳的角度上讲，牛顿所看到的光谱却有一个问题，那

就是它不仅可以从阳光中得到，而且也能从其他白色或接近白色的光源中得到，因此它带给我们的信息似乎并不是太阳所特有的。那么，在阳光里是否还隐藏着更微妙的信息，甚至是太阳所特有的信息呢？由于牛顿不能长生不老，对这些问题的探索就要依靠牛顿之后的科学家了。可惜的是，牛顿实在超前得太多了，在接下来一百多年的时间里，无论科学家们怎么重复和改进他的实验，都只能看见与他看见过的相同的彩色影像。这影像会不会就是大自然给我们的终极答案呢？没有人知道，但科学家们没有气馁，他们持续不断地进行着新的尝试。正是因为科学界有这样的恒心和毅力，像牛顿那样的高人也终有被超越的一天。

功夫不负有心人，在19世纪到来后的第二年，1802年，一位英国化学家终于窥视到了黎明前的曙光。所不同的是，这缕曙光不是彩色，而是黑色的！发现这缕曙光的英国化学家就是我们要介绍的第二位前辈，他曾经是一位医生，后来转向了科学，他的名字叫做沃拉斯顿（William Wollaston，1766—1828）。

1802年，沃拉斯顿对牛顿的实验进行了重复和改进。他采用了质量很好的三棱镜，并用狭缝取代了牛顿的小孔（以便让更多的阳光进入），结果他发现了一个前人不曾发现过的细节：在那熟悉的彩色光谱中，存在几条很细的暗线。那些暗线是什么呢？沃拉斯顿作了一个猜测，认为它们大都是不同颜色之间的分界线。他的这个猜测在当时听起来是有一定道理的，因为他所发现的暗线只有寥寥数条，而人们描述光谱所用的颜色也只有寥寥数种（红、橙、黄、绿、蓝、紫等），两者之间的确存在粗略的匹配性。但他没有想到的是，那些看似不起眼的暗线，已经让他站在了一座巨大冰山的尖顶上。11年后，当那座冰山的更多部分显露在人们面前时，他的猜测就不攻自破了。

那位让人们窥知冰山更多部分的人，就是夫琅禾费（彩图6）。

夫琅禾费出生于一个光学仪器世家，父亲和爷爷都是玻璃工匠，母亲那一方与玻璃工艺的渊源更是可以回溯到17世纪早期。但很不幸的是，在夫琅禾费10岁和11岁时，他的母亲和父亲先后去世，慕尼黑的贫民窟里从此多了一位年幼的孤儿。在随后的几年时间里，夫琅禾费靠替一位镜片制造商做学徒维持着艰难的生计。1801年，厄运再次降临到他的身上，他所住的贫民窟里的“脆脆楼”垮塌了。好在14岁的他被人从瓦砾堆下救了出来，算是不幸中的万幸。夫琅禾费的悲惨遭遇引起了恰巧途经垮塌现场，后来成为巴伐利亚国王的马克西米利安一世

（Maximilian I, 1756—1825）的同情，在他的资助下，夫琅禾费进了学校，并在十年后成为了光学研究的高手。1814年，夫琅禾费发明了光谱仪。这种仪器的核心部分虽仍是三棱镜，但在三棱镜的前后分别用透镜或透镜组对光线进行了汇聚，从而大大提高了分辨率。

有了光谱仪的帮助，隐藏在阳光里的真正奥秘终于比较清晰地显露在了人们面前。与沃拉斯顿一样，夫琅禾费也看到了暗线。但他的光谱仪远比沃拉斯顿的三棱镜精密，因此他看到的暗线不是寥寥数条，而是有几百条之多，他对它们进行了仔细的编号。为了纪念夫琅禾费的贡献，人们把那些暗线称为了夫琅禾费线（Fraunhofer lines）。后来随着光谱仪技术的进一步改良（比如使用更好的三棱镜，更精密的透镜，使用光栅等），以及照相技术的加盟，人们在阳光光谱中观测到的暗线数目也越来越多。

但那些暗线到底是什么呢？这个曾经困扰沃拉斯顿的问题也困扰着夫琅禾费。他首先怀疑的是自己的仪器：那些暗线会不会是自己仪器的缺陷造成的呢？他对这种可能性进行了排查，排查的方法很简单，那就是观察其他光源。如果暗线是仪器的缺陷造成的，那就应该与光源无关，从而应该同样地出现在其他光源的光谱中。观察的结果很快排除了

那种可能性——其他光源的光谱中并没有出现同样的暗线分布。这样，夫琅禾费就得到了一个结论：太阳光谱中的暗线是阳光本身的特征。由于那些暗线看上去虽然繁杂，却每一条都有固定的位置（这也是夫琅禾费能对它们进行编号的基础），它们显然隐藏着某种奥秘，而且这奥秘必定与太阳有关。

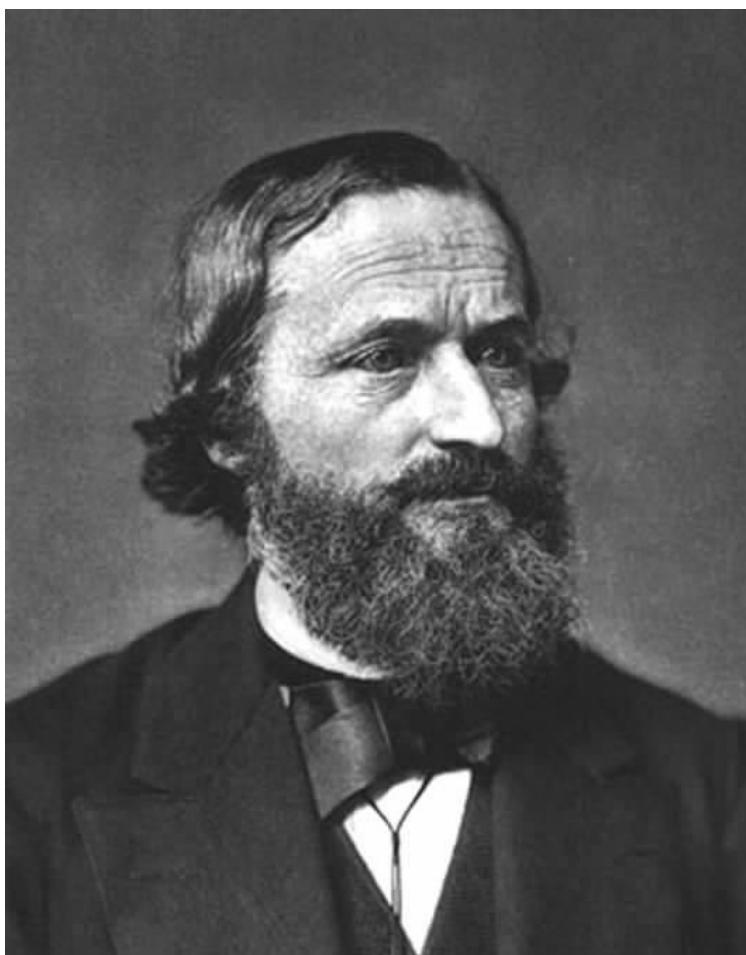
那么，这奥秘究竟是什么呢？夫琅禾费不知道，其他人也不知道。这局面多少有点尴尬，就好比已经发现了通往阳光奥秘的大门，却找不到开门的钥匙。这种尴尬局面持续了四十多年，在此期间，“功臣”夫琅禾费和沃拉斯顿，“反面陪衬”孔德等都先后离开了人世。

解密的日子终于到来了。

1859年，两位德国人在光谱研究上取得了突破性的成果。这两位德国人一位是化学家，名叫本生（Robert Bunsen, 1811—1899），以他名字命名的“本生灯”（Bunsen burner）直到今天仍被许多化学实验室所使用；另一位是物理学家，名叫基尔霍夫（Gustav Kirchhoff, 1824—1887），以他名字命名的“基尔霍夫电路定律”（Kirchhoff's circuit laws）直到今天仍是求解电路问题的重要工具。这两人当时都在海德堡大学（University of Heidelberg），本生当时正在研究化学元素被加热后所发射的光谱，那些光谱中有一些亮线，而且不同元素的亮线位置是不同的。本生打算利用这一特点作为证认化学元素的新手段。这在当时是一个很高明的想法。不过想法虽然高明，他用来观测光谱的设备却是滤色片一类老掉牙的东西，精度很低。这时候，他的朋友基尔霍夫给他支了一招，建议他使用光谱仪。本生采纳了这一建议。在接下来的一段时间里，本生和基尔霍夫进行了合作，他们不仅证实了每种化学元素都有自己独特的光谱，就像每个人都有自己独特的指纹一样，而且还通过光谱研究发现了两种新元素：铯（Cesium）和铷（Rubidium），显示了这

种手段的巨大威力。

在研究中，基尔霍夫自己也作出了一生中又一项重要发现。他注意到，如果某种元素在加热后所发射的光谱中有某些亮线，那么当光穿越由该元素制成的稀薄冷蒸气时，在光谱中原先的亮线位置上就会出现暗线。由于亮线源于光的发射（相应的光谱被称为发射光谱），暗线源于光的吸收（相应的光谱被称为吸收光谱），因此基尔霍夫的发现也可以表述为：一种元素能发射什么样的光，它也就能吸收什么样的光，两者相互对应。这个重要规律后来被称为基尔霍夫热辐射定律（Kirchhoff's law of thermal radiation）。^[3]现在我们知道，光谱线是电子在不同能级之间跃迁产生的：电子从高能级跃向低能级时会发射能量，由此产生的是发射光谱，从低能级跃向高能级时会吸收能量，由此产生的就是吸收光谱。由于这两者是由同一组电子能级决定的，它们之间相互对应也就不足为奇了。不过这一切直到20世纪初才随着量子理论的发展而被人们所了解。在基尔霍夫的时代，人们对光谱线的了解还停留在“知其然，却不知其所以然”的水平上，基尔霍夫热辐射定律也只是一个经验规律。



德国物理学家基尔霍夫（1824—1887）

但把这一经验规律与他和本生所做的事情联系起来，基尔霍夫立刻意识到了这一整套方法的重要价值。很明显，太阳光谱中的夫琅禾费线——那成百上千条的暗线——正是太阳上的吸收光谱，[\[4\]](#)既然吸收光谱与发射光谱是相互对应的，那么我们只要将夫琅禾费线与已知元素的发射光谱相比较，就可以证认出太阳上的元素。这样，基尔霍夫就找到了开启大门，破解阳光奥秘的钥匙。对于基尔霍夫的这一重要发现，本生曾经作过这样的记述：

基尔霍夫和我现在正忙于一种让我们夜不能寐的研究之中。基尔霍夫做出了一个有关太阳光谱中暗线起因的奇妙的、完全出人意

料的发现。他能在太阳或火焰的连续光谱中，在与夫琅禾费线严格对应的位置上，以人工加强的方式产生出那些暗线。这就找到了一种方法，能让我们像确定化学试剂中的氯化铯那样精确地确定太阳和星星的组成。

利用基尔霍夫的发现，科学家们很快就在太阳光谱中辨认出了大量和地球上相同的元素（孔德同学的悲观预言正式入住历史博物馆）。有些读者或许还记得，我们在第3章中曾经提到过一种叫做“天贵地贱”的古代观念，按照那种观念，完美的天体和卑微的地球是由完全不同的质料组成的。那种观念后来因伽利略发现太阳黑子而遭到了驳斥。不过，伽利略所驳斥的只是天体的完美性，对那种观念的真正毁灭性的打击，则是来自人们对天体化学成分（即所谓质料）的了解，因为是它直接证实了天体与地球由相同质料所组成，而这正是光谱学手段的功劳。

随着研究的深入，科学家们不再满足于像牛顿那样利用普通的日光来做研究，他们开始寻找新的机会，其中最重要的机会就是日全食。在1868年8月18日的日全食（属于沙罗序列133）期间，法国天文学家詹森（Pierre Janssen, 1824—1907）远赴印度观测了太阳光谱，结果在太阳大气层中的色球层（chromosphere）的光谱中发现了一条波长为5 875 Å的黄色亮线。^[5]不同寻常的是，这条亮线在地球上任何已知元素的发射光谱中都找不到对应。这是怎么回事呢？是詹森搞错了吗？有可能，但这种可能性很快就被排除了，因为两个月后，英国天文学家洛克耶（Norman Lockyer, 1836—1920）也在对同一次日全食的光谱记录中发现了同样的亮线。两组独立的观测同时搞错，而且错得一模一样的可能性无疑是很小的。

看来那条亮线并不是错误，既然不是错误，那它是什么呢？联想到

本生和基尔霍夫通过光谱学手段发现新元素的故事，这一问题的答案显然已呼之欲出。是的，答案就是新元素——一种在地球上尚未被发现过的新元素。由于这一元素是在太阳上发现的，洛克耶将之命名为“氦”（Helium），这个名称来自表示太阳的希腊语helios(ἥλιος)。不过，氦虽然是在太阳光谱中被发现的，它作为“太阳元素”的身份却并未维持很久。1895年，人们在地球上也发现了这种元素。光谱学方法不仅让我们了解了太阳的化学组成，而且还帮助我们发现了新元素。追根溯源，这一切当然都离不开夫琅禾费所发明的光谱仪，它是如此地卓有成效，以至于英国天文学家德拉鲁（Warren De la Rue, 1815—1889）曾经表示，即便我们能跑到太阳上去，把一些太阳上的物质拿到实验室来研究，也不会比用光谱仪得到的结果更精确。为了纪念夫琅禾费的重大贡献，人们在他的墓碑上刻下了这样一句墓志铭：他使星星变得更近。事实上，他不仅使星星变得更近，还为一个新领域的开创奠定了基础，因为整个天体物理学都是随着光谱学方法的应用而产生的。

氦元素的发现引起了人们对日全食太阳光谱的更大兴趣。第二年，1869年8月7日的日全食（属于沙罗序列143）期间，天文学家们又对太阳光谱进行了仔细观测，结果居然又有新的斩获：他们在比色球层更外部的所谓日冕（corona）的光谱中发现了一条波长为5 303 Å的绿色亮线。这条亮线在所有已知元素的发射光谱中也找不到对应。在随后的若干年里，这样的新发现一而再，再而三地出现，人们在日冕光谱中发现的新谱线居然增加到了24条之多。看来除了氦以外，太阳上还有其他新元素。至于那24条新谱线是来自一种还是多种新元素，暂时还没法知道，人们姑且先取了一个名字，叫做“氦”（coronium），它来自日冕的英文名corona。但与后来在地球上找到了的氦不同的是，那神秘的“氦”元素从未在地球上露过面。难道“天贵地贱”的古老观念毕竟还是有那么一丁点儿的正确，太阳上毕竟还是有一些地球上不存在的元素吗？这

个疑问直到20世纪30年代后才真相大白，原来并不存在什么“氦”元素，那些日冕光谱中的新谱线乃是来自一些被剥去了外层电子后的金属离子（比如铁离子、钙离子和镍离子）。以那条5 303 Å的绿色亮线为例，它是来自被剥去了13个电子（占总数的一半）的“无上装”铁离子（ Fe^{13+} ）的发射光谱。

现在我们知道了太阳上有哪些元素。但科学家们的胃口比这更大，他们不仅关心太阳上有哪些元素，而且还想知道它们各自所占的比例——用天文学家们的术语来说，叫做元素的丰度（abundance）。这对于进一步探索发生在太阳上的物理过程具有极大的重要性。幸运的是，物理学的威力完全能满足科学家们这一得寸进尺的要求。

推算元素的丰度从技术上讲并不容易，但基本原理却并不复杂。我们已经知道，谱线的位置可以用来确定元素的存在，但谱线带给我们的信息并不只是位置。当我们观测到一条谱线时，还会得到另一条信息，那就是谱线的强度。这条额外信息所携带的正是有关元素丰度的知识。为什么呢？因为无论吸收光谱还是发射光谱，它的强度都与产生谱线的元素的丰度有关，丰度越大，谱线就越强，反之，丰度越小，谱线则越弱。知道了这一关系，我们就可以用谱线的强度来反推元素的丰度。当然，这是典型的“站着说话不腰疼”，实际计算起来有许多技术性的困难需要克服，因此这种计算直到1925年才有人去做。



英国天文学家佩恩（1900—1979）

最早进行这种计算的是一位年仅25岁的女博士，名叫佩恩（Cecilia Payne, 1900—1979），出生于英国。1919年，当爱丁顿验证广义相对论的故事风靡世界的时候，她有幸聆听了爱丁顿的讲座，从此爱上了天文学。不过在英国这样一个社会习俗比较保守的国家里，女性从事天文研究是很困难的（其实别说从事天文研究，她所就读的剑桥大学当时甚至不给女性颁发学位），于是她远渡重洋到美国去读博士。在读博士期间，佩恩对太阳上的元素丰度进行了研究。1925年，她的研究得出了一个惊人的结果：这个太阳系里最重的天体竟然主要是由元素周期表上最轻的元素——氢和氦——组成的！虽然当时人们对太阳上的元素丰度还一无所知，但自从“天贵地贱”的观念破灭之后，很多人已经想当然地走

到了另一个极端，认为太阳和地球有着相似的组成。在这种情形下，佩恩也知道自己的结果有点骇人听闻，为保险起见，她在发表前特意征求了一下当时恒星光谱研究的权威，著名美国天文学家罗素（Henry Russell, 1877—1957）[\[6\]](#)的意见。

没想到这一征求征出了苦恼，因为罗素给了她一个很负面的回答，他表示佩恩的结果是“显然不可能的”。受罗素的威名所慑，佩恩修改了论文的措辞，表示自己所发现的氢和氦的丰度“被认为是有问题的”，“几乎可以肯定是不符合实际的”。这样的措辞很快就被证实为是谦虚得过火了。因为在接下来的几年间，其他天文学家们也陆续独立地得到了与佩恩一样的结果。四年之后，就连罗素本人也得到了同样的结果。在铁的事实面前，罗素终于意识到自己错了，在他的论文中，他宣布自己的结果与四年前佩恩的结果有着“很令人满意的一致”。权威毕竟是权威，四年前，因为罗素的影响，佩恩修改了措辞，弱化了结论，而四年后，同样也是因为罗素的影响，佩恩的结果加速成为了主流，她的博士论文更是被誉为了天文学领域中最重要 的博士论文。在那之后，佩恩继续从事着天文学研究。1934年，她与一位俄国天文学家结了婚。1956年，她成为了哈佛大学有史以来第一位女性正教授及系主任。

如今我们知道，太阳是一个巨大的“氢气球”，在它的总质量中，氢占了71%（在原子数目中则占了91.2%），氦占了27.1%（在原子数目中则占了8.7%），其余所有元素加起来也只占不到2%（在原子数目中则只占不到0.2%）。事实上，不仅太阳如此，绝大多数“青壮年”时期的恒星都是如此，甚至在目前整个宇宙的可见物质中，氢和氦也是绝对的“主流”。

在本章的最后，我们要对元素丰度的计算再作两点补充。细心的读

者也许注意到了一个问题，那就是太阳光谱——无论吸收光谱还是发射光谱——主要来自太阳大气层，因此通过光谱学研究得到的元素丰度应该是针对太阳大气层而非整个太阳的。事实的确如此。不过另一方面，太阳是一个高温气态的星球，托这种恶劣环境的福，在太阳内部很大的体积范围内都存在强烈的对流，使得太阳大气层的元素丰度与内部基本一致。不仅如此，对太阳模型及太阳起源的研究都显示，即便在不存在显著对流的区域，元素丰度仍与外部接近（唯一的例外是核心区）。因此，我们通过光谱学手段得到的元素丰度是对太阳整体元素丰度的一个不错的近似（确切地说是重元素丰度略微偏低）。

另一点需要补充的是，有关太阳元素丰度的计算需要用到统计物理、量子力学等领域的知识，且计算量相当巨大，很容易出错。这其中有一个著名的例子，那就是人们在铁元素丰度计算的量子力学部分中曾经犯过一个错误，导致铁元素的丰度被低估了90%，那个错误直到1968年才被纠正。除计算量巨大外，元素丰度计算所面临的另一类困难，是有些元素——比如氦——的谱线产生于色球和日冕中的一些远离平衡态的区域，从而很难建立理论模型。不过，我们今天所知的太阳元素丰度已经不单纯是光谱分析的结果，而是得到了一些其他手段，比如日震学（helioseismology）手段的印证，因此具有比单一手段更大的可靠性。

[1]当然，在几何规律的背后有着动力学的规律，比如牛顿运动定律和万有引力定律。不过在太阳系中，太阳是一个例外，它的质量远大于行星质量，具备“稳坐钓鱼台”的资格，因此在研究太阳时较少涉及动力学。

[2]牛顿的三棱镜实验是从1664年开始的，比较完整的实验则是在1666年初做的。在牛顿之前，也有学者注意到过类似的现象，并提出过

类似的想法（但都不如牛顿的细致和系统）。比如出生于爱尔兰的英国科学家玻意耳（Robert Boyle, 1627—1691）曾用三棱镜观察到过阳光里的彩色；牛顿的老冤家英国科学家胡克（Robert Hooke, 1635—1703）则曾经提出过颜色由蓝色与红色两种基色组成的想法（后来被牛顿推翻）。

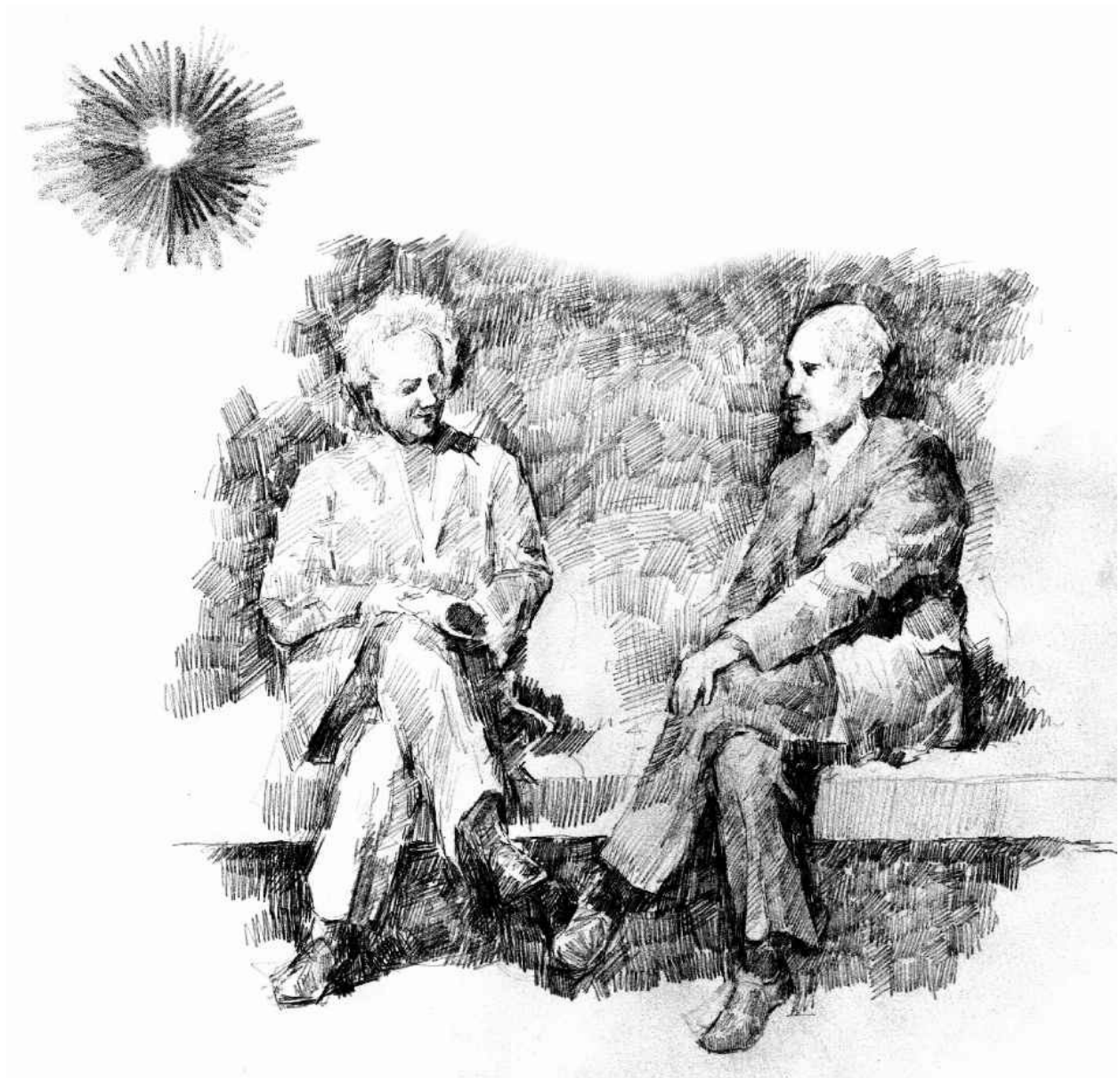
[3]我们在前面曾经提到过，为了判定太阳光谱中的暗线是否来自仪器缺陷，夫琅禾费曾检验过其他光源，他当时也注意到了太阳光谱中的某些暗线对应于普通光源中的亮线，可惜他没有对之作进一步的考察，从而错过了作出同样发现的机会。

[4]细心的读者也许会提出两个问题。一个是：基尔霍夫热辐射定律中的吸收光谱是由冷蒸气产生的，太阳上的温度那么高，为什么也有吸收光谱？答案是：所谓“冷蒸气”是相对于光源温度而言的。太阳吸收光谱主要来自太阳大气层中的光球层（photosphere），那里的温度用日常标准来衡量虽然极高，相对于太阳内层而言，却是一个“凉风习习”的地方，它的吸收作用大于发射作用，因而仍是“冷蒸气”。第二个是：阳光从太阳来到地球，除穿越太阳大气层外，还必须穿越地球大气层，那么，夫琅禾费线会不会是来自地球大气层呢？答案是：的确有少数夫琅禾费线是来自地球大气层的，比如被夫琅禾费标记为A和B的两条暗线实际上是来自地球大气层中的氧分子。在地球大气层之外观测太阳光谱时就不会看到那两条暗线。

[5] $1\text{\AA}=10^{-10}\text{m}$ 。读者也许会觉得奇怪，太阳的光谱是吸收光谱，光谱线不都是暗线吗？怎么忽然冒出一条亮线来了？这是有原因的，原因就是色球层中存在比内层温度更高的区域，使光谱由吸收光谱变成了发射光谱（由此可见，光谱不仅能帮我们辨认太阳上的元素，还能带给我

们有关太阳温度的信息)。除色球层夕卜,下文即将提到的日冕中的亮线也是如此。至于色球层与日冕这两个太阳大气层的外围区域为什么反而会有高温,我们将在第13章中加以介绍。

[6]读者请勿将这位罗素与哲学家罗素(Bertrand Russell, 1872—1970)混淆起来。此罗素非彼罗素。此罗素是一位天文学家,他一生最著名的工作是与丹麦天文学家赫茨普龙(Ejnar Hertzsprung, 1873—1967)各自独立地发现了恒星的光谱类型与光度之间的关系,即著名的赫罗图(Hertzsprung-Russell diagram)。



绘画：张京

7 物理自助游：推算太阳的质量、光度和表面温度

自第6章开始，我们的太阳故事已延伸到了物理领域。由于我们有关太阳的现代知识几乎全都来自物理手段，因此在本章及以后的多数章节中，我们将继续与物理“亲密接触”。说到物理，很多读者的脑海里也许会浮现出中学物理课上学过的一些基本概念，这其中很重要的一个就是被牛顿称为“物质的量”（quantity of matter），并在其名著《自然哲学的数学原理》的第一页上就试图定义的概念：质量。

这样一个重要概念当然也适用于太阳。本章的第一个任务，就是要查一查太阳的“资产”——它的“物质的量”。对于我们日常生活中接触到的普通物体来说，测定质量是一件毫不困难的事情。天平、杆秤、磅秤等都可以帮我们达到这一目的。这些测量手段有一个共同特点，那就是借助于所谓的静力学平衡手段。在历史上，早在质量的含义还仅仅停留在字面意义上那个“物质的量”，而不涉及像惯性和引力那样的动力学性质的年代里，人们就是用这类手段来测定质量的。

但是，就像大小和远近这样的简单测量一涉及太阳就变得不再轻而易举一样，质量的测定一涉及太阳，也就立刻变得棘手了。无论天平、杆秤还是磅秤，想用来测定太阳的质量，都无异于是白日做梦。怎么办呢？我们想起了第6章的思路。在第6章中我们介绍过，为了探索太阳的化学组成，科学家们研究了阳光，它给我们带来了远在1.5亿千米以外的太阳的信息。那么，除阳光外还有没有别的什么东西也能够不受遥远距离的阻隔，为我们带来有关太阳的信息呢？有，那就是引力，它不仅能够向我们申报太阳的“资产”，而且还采取了一种我们自己就能实践的方

式，为我们再次开展“自助游”活动提供了便利。

我们这次自助游所采用的将是物理手段，其中首先要利用的就是引力。

与上次自助游所采用的几何手段不同，人们对引力的了解要晚得多。如果我们想再玩一次重返古希腊之类的游戏的话，我们要重返的将不是古希腊，而是17世纪。另外，与阳光能够直接产生视觉不同，引力本身是看不见摸不着的。了解引力的主要途径，是研究它所导致的物体——尤其是天体——的运动，这正是17世纪科学家们曾经做过的事情。不过17世纪毕竟不同于古希腊。古希腊的很多推理在今天已是一些中小學生都能理解或反驳的，但17世纪的某些科学成就即便在今天，也足以难倒不少理工科的大学生。因此本章虽也号称“自助游”，却不能像上一次那样“徒步”进行，而需要先介绍一些17世纪的东西，作为代步工具。当然，这样做的另一个目的是让整本书的内容更加完整。

我们刚才提到，了解引力的主要途径，是研究它所导致的物体——尤其是天体——的运动。这种研究曾经导致了像日心说那样的重大天文进展。但对于了解引力本身而言，真正的进展却是出现在17世纪初期的1609年。那一年，一位德国天文学家发表了几项重大成果。此人的大名我们在第3章中已经提到过，他就是开普勒。是他，发现了行星运动的椭圆轨道，将沿用两千年的超级教条——天球——送进了历史博物馆；同样也是他（而不是哥白尼），使日心说在精度上超越了地心说。

但在作出这些辉煌成就的同时，他也亲手葬送了一个13年前（1596年）由他自己提出，且受他钟爱的模型。那个模型将当时已知的太阳系六大行星（水星、金星、地球、火星、木星、土星）的天球用三维空间中仅有的五种正多面体（正八面体、正二十面体、正十二面体、正四面

体、正六面体）以内切和外接的方式相分隔。那个模型不仅让开普勒一度以为发现了上帝创世计划中最宏伟的几何设计，而且还引起了当时已负盛名的丹麦天文学家第谷^[1]（Tycho Brahe, 1546—1601）的注意。1600年，开普勒应邀成为了后者的助手。一年后，第谷去世，他的职位及观测数据都被开普勒所继承。在望远镜时代来临之前，第谷的观测数据堪称举世无双。正是有那样精密的数据做后盾，加上自己的常年努力，开普勒才发现了他的行星运动定律，追根溯源，很多缘分都来自他当年那个模型。可惜再珍贵的模型如果与观测不符，也只能被放弃。假如科学研究本身也有定律的话，这或许就是第一定律，而开普勒很了解这一定律。

开普勒在1609年所发表的不仅有被称为开普勒第一定律的行星椭圆轨道，而且还有所谓的开普勒第二定律，即行星与太阳的连线在单位时间内扫过恒定的面积。但这两条定律加在一起，似乎也赶不上当年那个模型所具有的和谐与秩序。那个模型虽然被迫放弃了，开普勒却深信它所体现的和谐与秩序必定还有其他体现方式，为此他继续进行着不懈的探索。十年之后（1619年），他终于找到了一条新的规律。他为这一新规律撰写了一部新著作（图7.1），书名就叫《世界的和谐》（*Harmony of the Worlds*）。在那部著作中，他提出了自己的第三定律：行星轨道半长径（即椭圆轨道长轴长度的一半）的三次方与轨道周期的平方之比是一个常数。



图7.1 开普勒的名著《世界的和谐》

开普勒第三定律是一条非常漂亮的定律，任何人只要看一眼那些数据之间近乎完美的关联，就不难欣赏到它的美。开普勒本人对它也非常满意，称之为“和谐定律”（Harmonic Law）。开普勒的这些定律是牛顿之前人们在研究天体运动方面所达到的最高成就，为最终发现万有引力定律做出了重要铺垫。不过对于我们想要做的事情，即推算太阳的质量来说，它们还不够，因为其中还缺少一样东西，那就是质量这个主角。事实上，开普勒三大定律全都是运动学定律，只涉及时间和空间这样的运动学概念，而没有质量和引力那样的动力学概念。^[2]

为了推算太阳的质量，我们还得等一个人，一个将质量概念全面引进物理的人，一个能对引力作出数学描述的人，这个人就是牛顿。公元1687年，牛顿发表了巨著《自然哲学的数学原理》。在这部著作中，他完整地阐述了自己的三大运动定律及万有引力定律，为物理学的发展开辟了崭新的道路。用另一位科学巨匠爱因斯坦的话说，牛顿所发现的道路在他那个时代“是一位具有最高思维能力及创造力的人所能发现的唯一道路”。

沿着这条“唯一道路”，人们对开普勒三大定律有了更深入的了解，那三大定律不仅适用于行星绕太阳的运动，而且也适用于卫星绕行星的运动。那么，我们所关心的东西——质量——在哪里呢？就在开普勒第三定律中的那个常数——即轨道半长径的三次方与轨道周期的平方之比（以下简称“开普勒常数”）——里。利用牛顿的运动定律和万有引力定律可以很容易地证明，那个常数正比于中心天体的质量。^[3]

主角终于露面了！

既然质量就出现在开普勒常数中，那我们是不是就可以用开普勒第三定律来计算太阳质量了呢？很遗憾，答案暂时还是否定的。因为质量

在开普勒常数中的出现还捆绑了一个陌生的、来自牛顿万有引力定律的东西：万有引力常数。开普勒常数正比于中心天体的质量是不假，但在比例系数中却包含了万有引力常数。因此只有知道了万有引力常数的数值，才能真正推算天体的质量。但万有引力常数的数值是多少呢？很可惜，那在牛顿时代还是一个谜。

搞了半天，原来是空欢喜一场。质量虽然出现了，却“犹抱琵琶半遮面”。但即便如此，我们依然有一件事情可以做，那就是推算太阳质量与地球质量的比值。推算的方法很简单：既然开普勒常数正比于中心天体的质量，那就说明地球轨道（它的中心天体是太阳）的开普勒常数正比于太阳质量，而月球轨道（它的中心天体是地球）的开普勒常数正比于地球质量，它们的比值则等于太阳质量与地球质量之比。这个计算是我们现在就可以做的，那个暂时让人摸不清路数的万有引力常数在计算比值时会自动消去。这个计算所需的全部数据都已在第1、2两章中介绍过了，即：

- 地球的轨道半长径（很接近太阳离我们的平均距离）：150 000 000 千米
- 地球的轨道周期（恒星年）：365.24天
- 月球的轨道半长径（很接近月球离我们的平均距离）：384 400 千米
- 月球的轨道周期（恒星月）：27.3天

由此可以得到：

- 地球轨道的开普勒常数： 2.5×10^{19}
- 月球轨道的开普勒常数： 7.6×10^{13}

当然，这两者都是有量纲的。（请读者想一想，它们的量纲是什么？）不过这量纲跟万有引力常数一样，会在求比值时自动消去，因此不必理会。这两个常数的比值约为330 000。这样我们就得到了一个重要结果：太阳的质量约为地球质量的33万倍。将之与我们在第2章中已经得到的“太阳的直径约为地球直径的109倍”联系起来，我们立刻可以得出另一个重要结论，太阳的平均密度约为地球平均密度的1/4，这跟我们在第5章末尾介绍过的太阳上轻元素占很大比例的结果在定性上是一致的。[\[4\]](#)

现在我们的处境与第2章中曾经遇到过的相差无几了，即有关太阳的数字已经与有关地球的数字连在了一起。如果我们有办法知道地球的质量，就可以顺藤摸瓜地得到太阳的质量。但问题是，地球的质量虽然只有太阳质量的三十三万分之一，却同样是不能拿天平、杆秤、磅秤之类的工具来测量的。为了测定地球的质量，我们同样必须借助引力。而一旦涉及引力，万有引力常数就是一道绕不过去的坎。

既然绕不过去，就只好放手一搏了。那么，怎样才能知道万有引力常数的数值呢？很简单，那就是测定一对质量已知的物体之间的引力。这个答案对谁都不是秘密，但万有引力常数却对谁都是秘密，因为这个答案所提议的测定在当时是连牛顿也没法办到的；因为引力这个貌似强大、能把硕大无朋的天体玩得团团转的力量，其实却是自然界中最微弱的相互作用。它的强大纯粹来自“集体的力量”，因而只有在天文尺度上才是显著的。可是天文尺度上的物体——即天体——的质量却全都指着万有引力定律来测定，这就变成了一个先有鸡还是先有蛋的问题：要想知道天体的质量，首先得知道万有引力常数的数值；而要想知道万有引力常数的数值，又首先得知道某些天体的质量。

这个近乎死循环的局面直到一百多年后的1798年才被打开。那一年，英国科学家卡文迪许（Henry Cavendish, 1731—1810）通过一个极精巧的扭秤实验（图7.2），破天荒地直接测定了两个普通物体（它们的质量当然是已知的）之间极为微弱的引力。虽然卡文迪许的目的是测定地球的平均密度（由此也可以直接得到地球质量），但从他的数据中不难算出万有引力常数的数值为 6.75×10^{-11} （这是国际单位制下的数值，量纲留给读者自行推算）。自那以后很多其他人也对万有引力常数进行了测定。引力在日常尺度上的微弱性给所有这类测定设置了公平的障碍，从中反衬出的则是卡文迪许的高超技艺，因为他所达到的精度直到一个多世纪后才有人超越。对万有引力常数的测定直到今天依然是一件很困难的事情。截至2006年，国际科技数据委员会（CODATA）对这一常数的最佳推荐数值为 $6.674\ 28 \times 10^{-11}$ ，相对误差约为万分之一。在所有基本物理常数中，这样的精度就算不是最低，也是接近垫底的。不过太阳和地球这样的庞然大物毕竟不是每盎司都得斤斤计较的黄金珠宝，这样的精度对于测定它们的质量来说已是绰绰有余了。

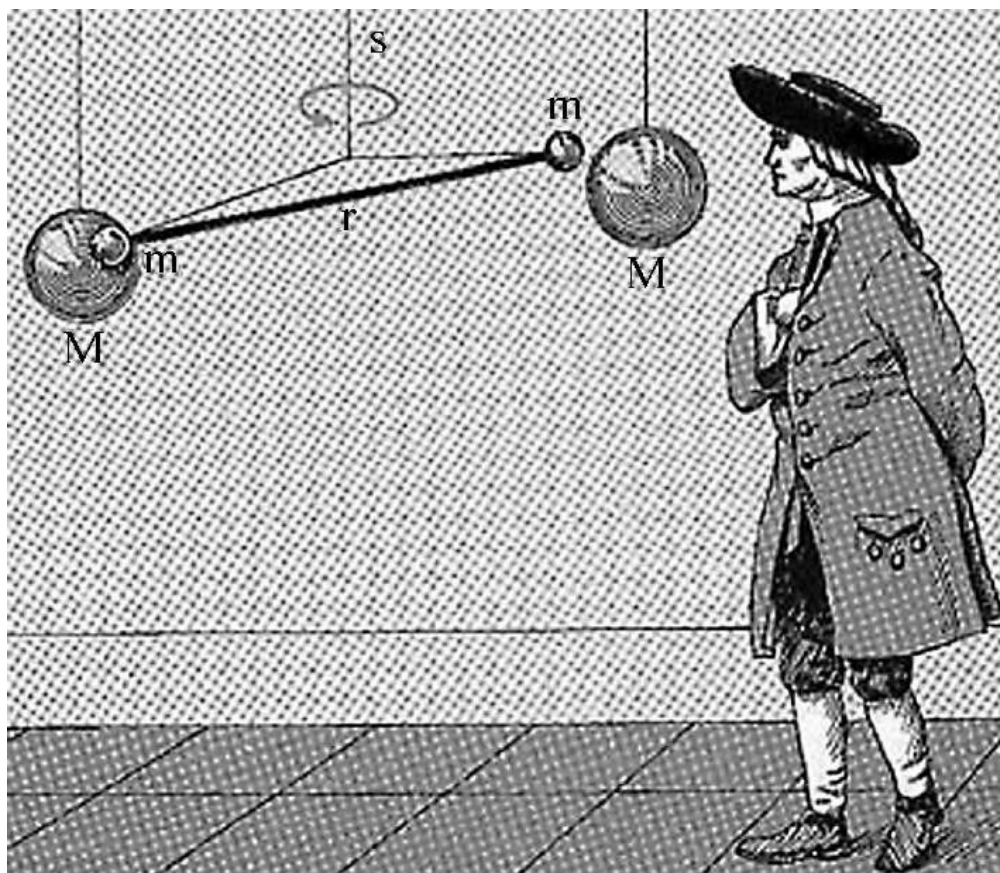


图7.2 卡文迪许的扭秤实验

知道了万有引力常数，我们就可以计算太阳和地球的质量了。计算的方法很多，既可以用开普勒第三定律，也可以通过其他方法，其中最简单的或许是将我们熟悉的地球表面重力加速度（ 9.8米/秒^2 ）与万有引力定律给出的加速度相对比。^[5]既然是自助游，这点小小的计算就留给读者们自己享用了，计算的结果将会表明：地球的质量约为 6×10^{24} 千克。^[6]将这一结果与前面已经推算出的太阳质量与地球质量的比值联系起来，就立刻可以得到太阳的质量约为 2×10^{30} 千克。

这样我们就完成了本次自助游的第一站：查明太阳的“资产”。“资产”既已查明，接下来我们就要去关心一下太阳的“开销”了，即它以电磁辐射（以下将笼统地称为“光”）的形式每秒挥霍掉的能量。天文学家

们把这种挥霍速度称为太阳的光度（luminosity）。

推算太阳光度的思路很简单，那就是测定地球公转轨道附近单位时间内垂直入射到单位面积上的阳光能量，即所谓的太阳常数（solar constant）。由于太阳很均匀地把光明洒向人间，因此一旦知道了太阳常数，将它乘上半径1.5亿千米（即地球公转轨道半径）的虚拟球面的面积，就可以得到太阳的光度。

那么怎样才能测定太阳常数呢？最简单的办法就是在阳光直射地面的时候，在地上放一盆“单位面积”的凉水，然后观察它在“单位时间”内的温度升高。将这一温度升高乘上水的比热和质量，就可以得到水从阳光中吸收的能量。在理想条件下，这个能量就等于太阳常数。当然，在实际实验中，阳光未必直射地面，水面面积未必是“单位面积”，观测时间也未必是“单位时间”，不过这些小小变通相信是难不倒读者的。

真正困难的是那“理想条件”四个字。要想切实做到这四个字，必须保证在那个“单位时间”里入射到“单位面积”上的阳光完全被水吸收，一点都不损失，而且还要保证那是水与外界唯一的能量交换。这两点要想切实做好显然都是很困难的。在历史上，最早测定太阳常数的是法国物理学家普耶特（Claude Pouillet, 1791—1868）和英国天文学家赫歇耳^[7]（John Herschel, 1792—1871）。1837年，他们两人彼此独立地进行了实验。“太阳常数”这一名称就是普耶特所取的。为了尽可能接近“理想条件”，普耶特（图7.3）和赫歇耳都做了一些努力，但效果并不理想，因为有一个因素在当时的条件下是无论如何也没法消除的，那就是地球大气对阳光的吸收和反射。这一因素不仅没法消除，甚至还会随时间、地点、阳光倾角、天气条件等因素的不同而不同，连校正都很困难。由于这种因素的干扰，普耶特测得的太阳常数只有正确值的一半左右，赫

歇耳的稍大些，但也强不了太多。

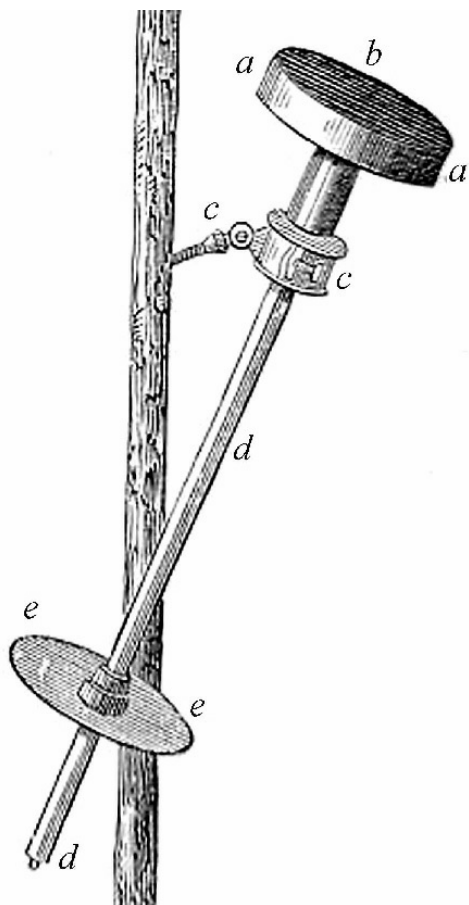


图7.3 普耶特测量太阳常数的仪器

（顺便请读者们思考一个“正大综艺”式的问题：仪器下方的圆盘e是做什么用的？）

为了减少大气干扰，天文学家们想了很多其他办法，比如将观测地点移到高山之巅。但要想真正摆脱地球大气的干扰，只有到外层空间去测量才行，这种“奢侈”的设想随着航天时代的来临成为了可能。从1978年到1998年的20年间，天文学家们利用人造卫星对太阳常数进行了持续测定，结果发现太阳常数约为 1366瓦/米^2 ，由此我们可以推算出太阳的光度约为 3.84×10^{26} 。 3.84×10^{26} 瓦是个什么概念呢？它相当于每秒钟爆炸920亿个百万吨级的氢弹！这样一个结果，居然可以从观察自家后院的

一盆水得到粗略的估计，这不禁让人想起一个有关美国物理学家费米（Enrico Fermi, 1901—1954）的故事来。1945年，在美国进行第一次核试验时，这位卓越的物理学家只用几张从空中飘落的纸片，就估算出了爆炸的当量（估算结果约为实际值的一半）。这种堪与福尔摩斯相媲美的奇妙推算无疑正是物理学的诸多魅力之一。

不过在这里，我们要提醒读者一点，那就是早在对太阳常数作出精确测定之前，人们就已意识到，所谓太阳常数其实不是一个真正的常数。即便扣除地球大气的干扰及地球离太阳的距离变化等因素，也没有任何物理理由表明太阳常数会是一个真正的常数。太阳常数的大小完全取决于发生在太阳上的物理过程。而再显而易见不过的事情就是，发生在像太阳那样的巨大星球上的物理过程是千变万化的，绝不可能给出一个简单的常数。事实上，对太阳常数的跟踪观测表明，它的数值随时都在变化，既有接近周期性的变化，也有非周期性的变化，不过变化的幅度倒是很小。

好了，现在让我们进入本次自助游的最后一站：推算太阳的表面温度（以下简称温度）。太阳很热是凡地球人都知道的事实，“赤日炎炎似火烧”嘛。但到底有多“热”呢，仅仅靠形容词是不够的，温度才是硬道理。因此很多天文学家都想知道太阳的温度。事实上，测定太阳常数的一个早期动机就是想推算太阳的温度。可惜的是，人们早期测定的太阳常数及由此推算出的太阳光度虽与现代值相距不远（起码在同一数量级上），但将它们与温度联系起来的理论基础却一直空缺着，这种理论“真空”导致了一片混乱的局面。拿推算温度的手法来说，可谓是五花八门，从子虚乌有的“热力线”（heat-ray）到并不适用的牛顿冷却定律（Newton's Law of Cooling），不一而足。拿推算结果来讲，从一千多度到几百万度，天差地别、应有尽有。为了鼓励可靠的研究，1876年，

法国巴黎科学院（Paris Academy of Sciences）特意为推算太阳温度设了一个奖，可惜还是无济于事，那奖被法国物理学家瓦耳勒（Jules Violle, 1841-1923）以一个很不靠谱的 $1\,500\sim 2\,500^{\circ}\text{C}$ 的推算所获得。

对太阳温度进行推算的理论基础直到1879和1884年，才先由奥地利物理学家斯特藩（Joseph Stefan, 1835—1893）从实验数据中得到，后由其同胞玻尔兹曼（Ludwig Boltzmann, 1844—1906）从热力学上推出。他们发现，一个黑体在单位面积上的辐射功率（即每秒钟辐射出的能量）正比于绝对温度的四次方。这一定律如今被称为斯特藩-玻尔兹曼定律（Stefan-Boltzmann law），其中的比例系数则被称为斯特藩-玻尔兹曼常数（Stefan-Boltzmann constant）。在国际单位制下，斯特藩-玻尔兹曼常数的数



奥地利物理学家斯特藩（1835—1893）

值为 5.67×10^{-8} （量纲仍留给读者自己去推导）。斯特藩-玻尔兹曼定律虽然针对的是黑体，但恒星辐射大都比较接近黑体辐射，因此该定律对恒星辐射也近似适用。有了这一定律，推算太阳的表面温度就有了理论基础。推算的方法很简单：（绝对温标下）表面温度的四次方乘上斯特藩-玻尔兹曼常数就是太阳表面每平方米的辐射功率，再乘上太阳的表面积，就是太阳的总辐射功率（即每秒钟辐射出的总能量），也就是我们前面刚刚推算过的太阳的光度。由此不难得到——请读者们“自助”完成——太阳的表面温度（确切地说是光球层的有效温度）约为 $5\,800\text{ K}$ （ K 为绝对温标的温度单位，摄氏温标的 0°C 约为 273 K ，本系列今后若提到温度而不注明温标，指的都是 K ）。

除上面这种方法外，我们再介绍一种虽然比较粗糙，但却别有趣味的方法。在上面的推算中，除用到斯特藩-玻尔兹曼定律外，还需要知道斯特藩-玻尔兹曼常数的数值，以及太阳的光度。但事实上，只要有斯特藩-玻尔兹曼定律所给出的四次方关系，即便不知道斯特藩-玻尔兹曼常数的大小，甚至不知道太阳的光度，我们依然能推算出太阳的表面温度。方法很简单：我们知道，地球表面的平均温度约为290 K（即17°C——这是对地域和时间的双重平均），虽然很容易被忽略，但这样温度的星球也会向外辐射能量，而且这个能量也可以近似地用斯特藩-玻尔兹曼定律来描述。由于不能像太阳那样自己发光，地球表面的能量主要来自阳光。^[8]因此地球能维持目前的表面温度，说明它向外辐射的能量与它所接收的阳光能量基本相等。利用这一关系，我们就可以推算出太阳的表面温度。推算的过程很简单，读者不妨自己试试，您将会发现，斯特藩-玻尔兹曼常数在推算过程中会自动消去，推算的结果约为6000 K，虽不如前面的方法精确，却也相差不远。

这个方法的趣味之处就在于它是利用行星的温度来反推恒星的温度。但更有趣的是，将它反过来用，我们也可以由恒星的温度来推算出一定距离外的行星温度。这一特点常被天文学家们用来估计恒星周围有可能抚育生命的所谓可栖息带（habitable zone）的位置和宽度。当然，这种推算的局限性是很大的，比如它既要求行星上存在水和大气那样能使温度均匀的东西，又要求那大气不能像金星大气那样富含温室效应气体。

自此我们就圆满完成了本次自助游的全部目标，即本章标题所宣称的推算太阳的质量、光度和表面温度。在“游戏结束”之前，我们再派发一个小红包——介绍一下太阳的光谱类型。在斯特藩-玻尔兹曼定律问世之后不久，德国物理学家维恩（Wilhelm Wien，1864—1928）用热力

学方法证明了一个定律，叫做维恩位移定律（Wien's displacement law）。这一定律表明，表面温度越高的物体，其光谱分布就越往短波方向偏移，表现在颜色上则是往蓝色方向偏移。利用这一特点，天文学家们将恒星依照光谱特征分为了七个大类，分别标记为O，B，A，F，G，K，M。其中O型天体为蓝色，表面温度最高，在33 000 K以上，M型天体为红色，表面温度最低，在3 700 K以下。像太阳这样的黄色天体为G型，表面温度在5 200~6 000 K之间。在每个类型之中，依照温度从高到低的顺序又分出十个亚型，分别用阿拉伯数字0~9来表示，0表示温度最高，9表示温度最低。太阳的光谱类型为G2，在G型之中算是温度较高的。^[9]

^[1]按照今天的称呼惯例，第谷·布拉赫应被称为“布拉赫”，就像艾萨克·牛顿被称为“牛顿”，阿尔伯特·爱因斯坦被称为“爱因斯坦”一样。不过在第谷所生活的年代，当地的习惯是用名而非姓作为称呼，因此第谷·布拉赫就被称为了“第谷”，并沿用至今。

^[2]在开普勒时代，引力的概念尚未形成。不过开普勒本人倒是很早就猜测过太阳是行星运动的原因。开普勒第三定律的发现使他更坚定了自己的猜测。因为这一定律表明，离太阳越近的行星运动得越快，这种相关性强烈地暗示着行星运动的原因来自太阳。据说开普勒是作出这种猜测的第一人（不过他所猜测的太阳对行星的影响类似于磁力，而不是引力）。

^[3]感兴趣的读者可以针对圆周运动这一特例自行证明这一点。当然，就像很多其他经验规律一样，开普勒定律的成立是有条件的，比如要忽略来自其他天体的影响。而且，开普勒第三定律中的那个常数实际上是正比于中心天体及绕其运动的天体的总质量，只有在中心天体的质

量远大于绕其运动的天体的质量时，才可以近似为中心天体的质量。对太阳系的全部行星和卫星而言，这一近似都基本成立。

[4]我们在后文中将会看到，太阳物质的密度分布是极不均匀的，从核心到外层有着天壤之别。

[5]不过这个号称简单的方法需要用到万有引力定律的一个不太简单的推论：那就是质量呈球对称分布的物体在其外部产生的引力相当于全部质量都集中在球心。这个推论在牛顿那个时代是只有牛顿才能证明的独门绝活，直到今天也足以难倒一部分理工科大学生。

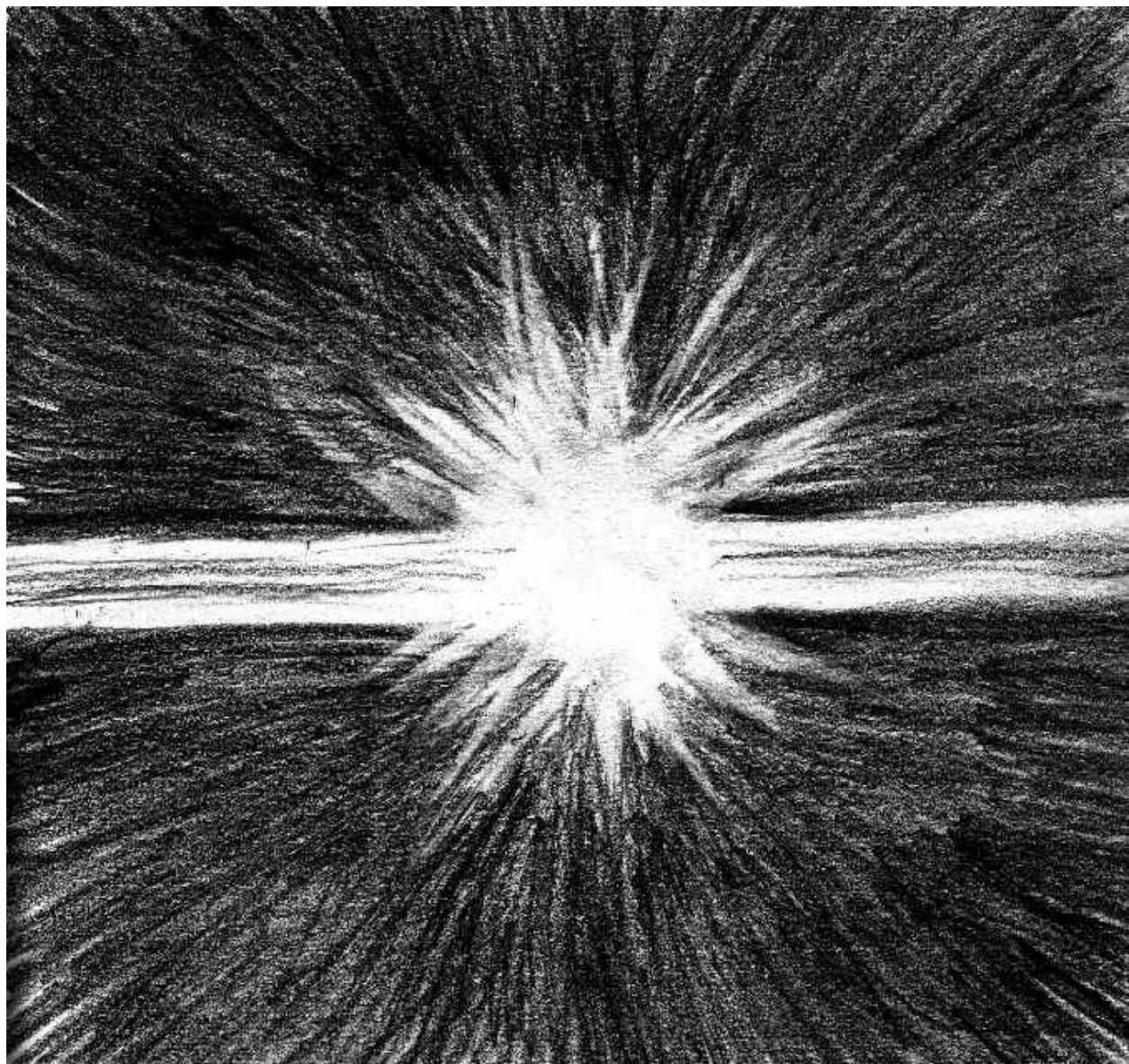
[6]将这一结果与地球的直径联系起来，就可以得到地球的平均密度约为 5.5克/厘米^3 。若进一步与前面得到的太阳平均密度约为地球平均密度的 $1/4$ 联系起来，则可以得到太阳的平均密度约为 1.4克/厘米^3 。

[7]这位赫歇耳通常称为小赫歇耳，他是天王星的发现者、有时称为老赫歇耳的威廉·赫歇耳（William Herschel, 1738—1822）的儿子。说到老赫歇耳，顺便提一下，他也在某种程度上测定过太阳的辐射能量。在1800年所进行的一组观测中，他曾将一个温度计放在太阳光谱的红端以外（他的本意是让那个温度计不被阳光照到，从而可以用来记录环境温度），结果发现记录下来的温度存在反常升高，由此他推断出阳光中存在着波长比红光更长的肉眼无法看见的部分，即红外线。

[8]除阳光外，在地球表面所接受的能量中也有来自地球内部的贡献。不过后者的平均功率只有 0.06瓦/米^2 ，不到阳光能量的万分之一，在我们的计算中可以忽略。

[9]在有些文献中，太阳的光谱类型被标记为G2V，这里的“V”所

表示的不是英文字母“V”，而是罗马数字的“5”，它来自另一种分类细则，所表示的含义是主序星（即壮年的恒星）。



绘画：张京

8 光明的源泉 恐怖的核心

爱因斯坦和他的年轻合作者英费尔德（Leopold Infeld, 1898—1968）曾经写过一本非常出色的科普著作，叫做《物理学的进化》（*The Evolution of Physics*）。在那部著作中，他们作过一个令人印象深刻的比喻，那就是把科学的发展比喻成一个侦探故事。他们这样写道：

自从柯南·道尔写出绝妙的故事以来，在几乎所有侦探小说中都有这样一个时刻，侦探收集到了为解决问题的某个阶段所需的全部事实。那些事实往往看起来很奇特、不连贯，并且彼此毫不相干。可是大侦探知道这时不必继续调查了，现在只有纯粹的思维能把搜集到的事实联系起来。于是他拉拉小提琴，或躺在安乐椅上抽抽烟。突然间，老天爷，他找到了联系！他不仅对手头所有的线索都有了解释，而且知道某些其他事情也一定发生了。因为现在他已经确切地知道在哪里可以找到它，如果愿意的话，他可以出去收集他理论的进一步证实。

我们的太阳故事进行到这里，也差不多到了那样一个时刻，可以——甚至必须——像大侦探一样解决一些谜团了。不过与大侦探的单枪匹马及“老天爷，他找到了联系！”那样的戏剧性不同，科学家们收集事实的过程更像是现代侦探连续剧中那些跨省市的联合行动。因为除了大侦探自己收集的事实外，来自同事、同行乃至其他部门的协助往往也对案情的侦破起着关键作用。而科学家们解决谜团的过程，与其说是“拉拉小提琴，或躺在安乐椅上抽抽烟”，不如说更像是现代侦探连续剧中的大型案情分析会。在那样的分析会上，很多人（尤其是新手）会提出很多设想。当那些设想被一一排除后，某位幸运儿（多半是男主人公）

会提出一个真正的好想法。当然，一个精彩的侦探故事往往是曲折的，一个曾经的好想法也许后来又遇到新问题，然后不得不召开新的案情分析会。我们将会看到，科学的发展常常也是如此。

现在就让我们来观摩一场有关太阳的案情分析会，会议的主题是：太阳为什么会发光？或者用更学术一点的话说，恒星的能量产生机制是什么？

太阳为什么会发光？这是一个很古老的问题。既然是很古老的问题，当然曾经有过很古老的回答，比方说“上帝安排的”。但那样的回答就好比是说凶案是上帝做的，不必查了。那样的人做神父远比做侦探合适，我们就不邀请了。[\[1\]](#)

那么，我们所邀请的第一位侦探是谁呢？是英国天文学家赫歇耳（William Herschel, 1738—1822）。此人是太阳系第六大行星——天王星——的发现者，在天文史上是一位赫赫有名的前辈。不过我们这场案情分析会虽然打算开成一个“团结的大会”、“奋进的大会”，却不是一个论资排辈的地方。在这个分析会上，赫歇耳的资格虽老，却必须被列入“菜鸟”行列（这也比较符合侦探片中新手总是先发言的惯例），因为他的提议确实是太“菜”了一点。那提议是这样的：太阳之所以发光，是因为它有一个因炽热而发光的大气层，那大气层的下面则有可能是一个凉爽、甚至可能有生命居住的固态表面！

今天的读者也许很难理解，像赫歇耳那样著名的天文学家怎么会想出如此不靠谱的提议？在这里，我们要为赫歇耳说几句公道话，因为他的提议在如今看来虽然极度离奇，在当时却并非单纯的臆想。事实上，赫歇耳之所以作出那样的提议，是因为他注意到了太阳上的黑子，他认为那是透过太阳大气层中的空隙所看到的太阳表面，那表面既然呈现黑

色，想必是凉的，这就是他那提议的依据。那依据虽然是错误的，在当时却不失为是一种可能性。在科学史上，错误的假说可以说是层出不穷，一些反科学人士往往以此为由来抨击科学。其实，科学之所以有今天的声望，从来就不是因为她不会出错，而是因为她不断地寻求实证，并且在这过程之中不断地纠正错误，去芜存菁。我们在后文中将会看到，黑子的“黑”是一种相对的颜色，它并不等于日常意义上的“凉”。我们在第6章中介绍过的太阳光谱的类型也间接证实了太阳“表面”非但不凉，而且要比大气层（确切地说是光球层）更热，因为否则的话，太阳大气层的光谱将会是发射光谱而非吸收光谱（参阅第6章的注释）。除上述误判外，赫歇耳的提议还有一个很大的问题，那就是无论多热的大气层如果得不到能量的补充，都会很快冷却下来，这与太阳长时间稳定地发光是完全矛盾的。^[2]

有这么多严重问题，赫歇耳的资格再老，也只能到一旁凉快去了。接下来有两人几乎同时发了言，这两人一位是德国医生兼物理学家迈耶（Julius von Mayer, 1814—1878），另一位是苏格兰物理学家沃特斯顿（John Waterston, 1811—1883）。说来也巧，这两人不仅发言时间相近，学术经历也相似：他们都在热力学领域中做过一些先驱性的工作，前者研究了热功当量和能量守恒，后者研究了气体分子运动论，但两人的工作都在很大程度上被同时代人所忽略。到了19世纪40年代后期，这对“难兄难弟”又几乎同时对太阳为什么会发光产生了兴趣。与赫歇耳时代不同，当这两人开始研究这一问题时，能量守恒的观念已初步形成（如前所述，迈耶本人就是研究这一观念的先驱之一），因此他们不能像赫歇耳那样随意假设一种近乎静态的发光机制，而必须寻求能让太阳持续发光的能量来源。

作为起点，他们两人都分析了一种土得掉渣的假说，即把太阳当成

一个燃烧的大煤球。这种“煤球说”的提出是不需要想象力的，凡用过煤炉的人都有可能想到，但推翻它却需要一些专业知识。经过计算，迈耶发现太阳虽然大得惊人，但与它那更惊人的光度相比却还不够大，假如太阳果真是个大煤球的话，即便维持燃烧所需的氧气不是问题，它也只能燃烧几千年。沃特斯顿的研究则表明，不仅大煤球烧不了多久，其他化学反应也强不到哪里去，至多能撑20 000年（感兴趣的读者可以利用化学过程所涉及的能量与原子外层电子的能量同一量级这一特点核实一下他们的结果）。如果时光能倒转两百年，他们这些结果将不仅不是问题，反而是天大的优点，因为那跟神学家们“推算”出的世界历史的长度大致相当。但在19世纪中叶，学术界已基本断定地球的年龄远不止那个量级。而依照当时流行的康德-拉普拉斯星云假说（Kant-Laplace nebular hypothesis），太阳和地球是由同一片星云收缩而成的，彼此年龄相近。既然太阳和地球的年龄相近，而地球的年龄远不止几千年，那么只能燃烧几千年的“大煤球”显然不可能是太阳持续发光的能量来源。因此“煤球说”也被排除掉了。

排除了“煤球说”之后，迈耶和沃特斯顿各自提出了自己的假说。迈耶提出太阳之所以持续发光，是因为不断有陨星坠落到太阳上。沃特斯顿则认为是太阳自身引力收缩产生的热量使它持续发光。两者之中，沃特斯顿的“引力说”由于与康德-拉普拉斯星云假说存在共性——即都与引力收缩有关——而占有一定优势。不过优势归优势，两人的论文投寄出去后的命运却是相同的：那就是都被拒掉了——迈耶的论文被巴黎科学院（Paris Academy）所拒，沃特斯顿的则被伦敦皇家学会（Royal Society of London）所拒。但即便如此，沃特斯顿还是找到机会于1853年向英国科学进步协会（British Association for the Advancement of Science）报告了自己的理论，他的报告打动了两位著名人物：德国物理学家亥姆霍兹（Hermann von Helmholtz, 1821—1894）和英国物理学家

汤姆孙（William Thomson, 1824—1907），即后来的开尔文勋爵（Lord Kelvin）。他们成为了我们这个案情分析会的下两位发言者。

这两人之中，亥姆霍兹的发言很简短，中心思想就是坚决拥护沃特斯顿的“引力说”。汤姆孙则比较健谈，他虽然也表态支持“引力说”，却详细回顾了自己的“心路历程”。他坦承，自己曾经喜欢过迈耶的“陨星说”，但在接触过程中逐渐认清了后者存在的问题，那就是假如太阳系中仍有那么多陨星，我们地球也应该会分到一杯羹，这与地球目前的“冷清”状态显然不符（事实上，考虑到我们这些地球生物的脆弱性，地球若果真分到那“一杯羹”的话，我们早就挂掉了）。不仅如此，假如太阳的巨大能量真的来自陨星，那陨星的数量必须极为庞大，它们的坠落将使太阳的质量增加，进而影响行星的轨道。由此导致的后果之一，是过去几千年中地球的公转周期应该缩短几个星期。这与天文观测显然也是矛盾的。“吾爱‘陨星说’，但吾更爱真理”，面对如此严重的问题，汤姆孙毅然决然地抛弃了“陨星说”，转而投入“引力说”的怀抱。

与“陨星说”相比，“引力说”的确显得高出一筹。亥姆霍兹和汤姆孙的研究表明，太阳的半径只要每年收缩几十米，就足以维持目前的光度（感兴趣的读者不妨自己估算一下，看能否证实他们的结果）。相对于139万千米的太阳直径而言，那样小幅的收缩在当时是任何人都无法察觉的，因而不与观测相矛盾。而且，那样的收缩可以持续几千万年，与汤姆孙本人所估算的地球年龄具有相同的量级。这一点给了他很大的信心，使他在有生之年里保持了对“引力说”的从一而终。若干年后，当人们利用新发现的放射性现象对地球年龄作了重新估算，发现它远比几千万年更古老时，汤姆孙依然固执己见，凭借自己的巨大威望将“引力说”全面推向了20世纪。

但在有关地球年龄越来越铁的证据面前，汤姆孙的威望虽高，终究

只是螳臂挡车。在这一点上，亥姆霍兹醒悟得比较早，他曾表示，如果能发现新能源，我们就可以把太阳的年龄延长。这虽然是一句没什么技术含量的大白话，却也道出了一个努力方向，那就是寻找新能源。只不过“煤球说”所用的化学能（本质上是电磁能）和“引力说”所用的引力能都被排除了，新能源在哪里呢？这个问题自然而然就成为了下一位发言者的主题。

这位发言者是我们的老朋友了，他就是英国天文学家爱丁顿，他的准备工作比前几位发言者都更充分，因为他从兄弟单位那里拉来了以下几项“友情赞助”（赞助者全都是诺贝尔奖得主）：

- 1905年，爱因斯坦提出了著名的质能关系式： $E=mc^2$ 。它表明在貌似寻常的物质之中蕴藏着惊人的能量。
- 1919年，新西兰物理学家卢瑟福（Ernest Rutherford, 1871—1937）在卡文迪许实验室里实现了所谓的人工原子核嬗变（nuclear transmutation），即用人工手段将一种原子核变成另一种原子核。^[3]
- 1920年，英国化学家阿斯顿（Francis William Aston, 1877—1945）发现了氢原子核（即质子）的质量要比重元素单个核子的平均质量略大。

这几项“赞助”落到爱丁顿那训练有素的大脑里，很快就变成了一个有关新能源的大胆构想：既然原子核可以彼此转变，而质子的质量要比重元素单个核子的平均质量略大，那么只要能把质子聚合成重原子核——比如氦核。前者所包含的多余质量就应该会按照爱因斯坦质能关系式所确定的“汇率”转变为巨大的能量。^[4]简单的估算表明，这种被称为核聚变（nuclear fusion）的能量产生机制足以支撑太阳100亿年以上，与

有关地球年龄的所有测定都完全相容。^[5]

当然，这只是理论可能性，它能否成为现实，关键得看太阳上的质子是否真能聚合成重原子核。在爱丁顿时代，那还是一个谜，因为核子世界的一个重要成员——中子——尚未被发现，有关核子相互作用的理论也尚未建立起来。不过爱丁顿很乐观，他表示“在卡文迪许实验室里能够做到的事情对于太阳来说应该不会困难”。但他乐观不等于别人也乐观，他的观点一经提出，就遭到了一位重量级英国物理学家金斯（James Jeans, 1877—1946）的强力反对。两人互不相让，展开了公开而激烈的争论，他们的争论一度成为皇家学会的一道风景，吸引了很多听不懂他们争论的科学家来看热闹（科学家也是人，他们的好奇心也并不总是集中在大自然上的）。

那么，金斯反对爱丁顿的理由是什么呢？最核心的一条理由是这样的：质子之间存在很强的静电斥力，为了让它们彼此接近到能够发生核聚变的程度，它们的热运动能量必须大到能克服静电斥力的程度，这要求极高的温度，而太阳内部是不可能那样高的温度的。对此，爱丁顿的回答是：像氦核那样的重原子核的存在是一个既有事实，如果连恒星内部的温度都不够高，那么宇宙中的重原子核从何而来呢？“我们不跟宣称恒星不够热的人争论，我们请他去找一个更热的地方”——他用这样一句掷地有声的名言把球扔回给了金斯。

这球金斯没能接住，因为当时没有人知道更热的地方（不过富有戏剧性的是，后来人们发现，还真的存在一个比恒星内部更热的地方，而且包括氦核在内的某几种重原子核确有相当一部分是在那里被产生出来的。那个地方就是大爆炸初期的宇宙）。爱丁顿虽然把球扔回给了金斯，但要想让别人真正信服，光靠扔皮球是不行的，他必须正面论述自

已理论的可行性。为此，爱丁顿对恒星结构模型进行了研究。在那些具有开创意义的研究中，他估算出太阳核心的温度约为4 000万度，核心物质的密度则为80克/厘米³（相当于黄金密度的四倍）。由于缺乏核相互作用理论的引导，他无法对核聚变的细节做出可靠描述，但他的估算结果在数量级上是大致成立的。^[6]

几千万度的高温，比黄金还高得多的密度，那样的数据看上去有些离奇。但福尔摩斯说得好：当你排除了所有的不可能，剩下的无论看起来多么不可能，一定就是真相。爱丁顿的理论就颇有那样的意味。如果我们对已被排除掉的“煤球说”、“陨星说”和“引力说”分一下类的话，那么从尺度上讲，它们中既有宏观的，也有原子尺度的；从相互作用上讲，则既有电磁的，也有引力的。既然那些都被排除掉了，剩下的尺度就该轮到原子核尺度，而剩下的相互作用则该是弱相互作用和强相互作用了，这正是爱丁顿假说的基本特点。^[7]但福尔摩斯的话虽然精彩，毕竟不是金科玉律。爱丁顿的假说真的代表真相吗？这个悬念我们要请下一位发言者来揭晓。那位发言者也做了充分准备，而且拉“赞助”的功夫也不含糊，总计拉到四项之多（赞助者除伽莫夫外，也全都是诺贝尔奖得主）：

- 1928年，俄国物理学家伽莫夫（George Gamow, 1904—1968）发现了量子力学的隧道效应，即微观粒子有一定概率穿越经典意义上不可穿越的“障碍”。这一发现在很大程度上破解了金斯的诘难，因为即便太阳核心不够热，依然有一部分质子可以通过隧道效应来克服静电斥力造成的“障碍”。
- 1932年，英国物理学家查德维克（James Chadwick, 1891—1974）发现了中子，为理解原子核结构铺平了道路。^[8]
- 1934年，意大利物理学家费米（Enrico Fermi, 1901—1954）提出

弱相互作用的四费米子理论（four-fermion theory），为近似描述核子反应中的弱相互作用部分提供了理论基础。

- 1935年，日本物理学家汤川秀树（Hideki Yukawa, 1907—1981）提出了强相互作用的介子理论，为近似描述核子反应中的强相互作用部分提供了理论基础。

有了上述“赞助”，我们的发言者很快就拟定了一份出色的发言稿，为爱丁顿的假说提供了决定性的支持。这位发言者是20世纪30年代从纳粹德国逃往美国的许许多多物理学家中的一员，他的名字叫做贝特（Hans Bethe, 1906-2005）。在贝特开始研究恒星能量产生机制时，对这一问题感兴趣的物理学家已经不少，其中包括后来的美国原子弹之父奥本海默（Robert Oppenheimer, 1904—1967）、美国氢弹之父泰勒（Edward Teller, 1908—2003），以及著名的苏联物理学家朗道（Lev Landau, 1908—1968）。不过贝特并不是一个人在战斗，他有合作者，此人名叫克里奇菲尔德（Charles Critchfield, 1910—1994），是一位研究生，在与贝特合作前曾跟随伽莫夫学习。

1938年，在伽莫夫的建议下，克里奇菲尔德研究了质子与质子之间的核反应。在得知贝特也在从事类似的研究后，伽莫夫让克里奇菲尔德把论文寄给贝特，由此促成了两人的合作。贝特和克里奇菲尔德的合作揭示了发生在太阳内部最重要的核聚变反应，即所谓“质子-质子链”（proton-proton chain, 简称pp链）的具体实现方式。这种实现方式中最主要的一类被称为第一类质子-质子链（简称ppI链），如彩图7所示。它是这样进行的：

（1）两个质子p聚合成氢的同位素氘核 ^2H （左上角数字表示核子数，下同）。

(2) 一个氘核 ^2H 与一个质子 p 聚合成氦的同位素 ^3He 。

(3) 两个 ^3He 通过扔掉两个多余质子 p 而聚合成一个标准氦核 ^4He 。

这整个过程释放出的能量约为26.7 MeV（MeV为百万电子伏特，是描述核反应能量的常用单位，约合 1.6×10^{-13} 焦耳），其效率是普通化学反应的几百万倍以上。在太阳核心所产生的能量中，这类质子-质子链的贡献占了约85%，是当之无愧的产能大户。^[9]

除质子-质子链外，贝特还研究提出了另一种恒星核反应机制，称为碳氮氧循环（CNO cycle），也称为碳氮循环（CN cycle），因为参与该反应的氧核是 ^{15}O ，而不是最常见的 ^{16}O 。这种反应所需的温度比质子-质子链更高，在像太阳这样的小质量恒星的能量产出中只占了1%左右。但它在质量比太阳大30%以上的恒星中却占据着主导作用，因而在恒星核物理中的总体重要性不亚于质子-质子链。与贝特几乎同时，德国物理学家魏茨泽克（Carl von Weizsäcker, 1912—2007）也独立地完成了同样的工作。魏茨泽克也是当时这一领域的主要研究者之一，并且正是受到他的影响，贝特才将自己的研究方向由单纯的核物理转到恒星核物理上。1939年，贝特写了一篇综述性的论文，对恒星能量产生机制作了比较完整的阐述。也许是由于对质子-质子链和碳氮循环这两类反应同时作出重要贡献的缘故，贝特成为了1967年度诺贝尔物理学奖的唯一得奖人，得奖理由中很重要的一条就是发现恒星能量产生机制。

在研究恒星能量产生机制的同时，人们对爱丁顿的太阳模型也进行了修正。其中最重要的修正是关于核心温度的，那温度虽仍是一个很难用日常指标来衡量的恐怖数字，但因为隧道效应的帮助，比爱丁顿所估计的4 000万度要低得多，约为1 570万度。这个温度的重估是十分必

要的，因为核聚变反应的剧烈程度与温度有着极为敏感的依赖性：温度越高，核聚变反应越剧烈。如果太阳的核心温度果真有4 000万度，整个太阳将会像一个超级氢弹一样，在惊天大爆炸中化为灰烬（我们的命运自然也可想而知）。与核心温度的调低相反，重新估计后的太阳核心密度则比爱丁顿的估计更高，约为160克/厘米³，而太阳的核心压强也极为惊人，达到2 500亿个大气压。这些数字的可怕之处一般人可能难以体会，我们可以举两个例子：维持在太阳核心温度上的物质，哪怕只有玻璃弹珠那样一小块，也足以熔化几百千米以外的钢铁；[\[10\]](#)而太阳的核心压强，则相当于在手指甲那样的面积上压上几亿吨的重物。可以毫不夸张地说，太阳核心这个太阳系的光明源泉，是一个不折不扣的恐怖核心。

好了，现在我们终于有了一个理论，像大侦探的理论一样，能对手头的线索（比如太阳的光度及维持这一光度的漫长时间）作出解释。从某种意义上讲，爱丁顿和贝特是幸运的，因为当爱丁顿开始思考恒星能量产生机制时，人们对物理世界的了解已经深入到了理解这一问题所必需的原子核尺度上，爱因斯坦也已经提出了质能关系式；而当贝特开始研究恒星能量产生机制的细节时，人们已经有了关于核子相互作用的初步理论。那些理论若是不存在，爱丁顿和贝特就是再高明十倍，恐怕也将会“巧妇难为无米之炊”。我有时会想，如果某个遥远星球上存在着其他智慧生物，他们会如何理解这个世界，我们称之为科学的东西在那里会以一种什么样的顺序发展呢？那是我心中有他们的最大好奇。至于他们长什么样？能活多少岁？对我来说倒是细枝末节。

最后让我们再回到本章开头所引的爱因斯坦和英费尔德的文字中来。有读者可能会问：爱丁顿和贝特等人的理论虽然给出了太阳的能量产生机制，但那恐怖核心却是一个我们永远也不可能到达的地方，我们

有什么办法来检验他们的理论呢？答案是：通过检验理论的推论。事实上，我们很快将会看到，由爱丁顿和贝特等人提出的这个恒星能量产生机制的确很像爱因斯坦和英费尔德所说的大侦探的理论，它不仅解释了我们手头的线索，而且也预言了某些其他事情一定已经发生。我们下一步要做的，就是去验证那些事情，收集有关我们理论的进一步证据。

但出乎意料的是，我们信心满满的收集行动却遇到了非常棘手的困难。大自然似乎嫌我们这个侦探故事还不够离奇曲折，而决心要为我们增添一些新的情节。在第9章中，我们将一同去面对那些困难，同时欣赏一下大自然为我们增添的新情节。

[1] 值得一提的是，在很古老的回答中也有相对不错的。比如公元前5世纪的古希腊先贤阿那克萨哥拉（Anaxagoras，公元前500—前428）曾经设想太阳是一个烧红的大铁球。理由是他亲眼目睹过如火球般从天而降的陨铁，他认为那是太阳的碎片。他这个提议当然也很“菜”，但起码与赫歇耳的提议有一拼。

[2] 不过这一点也不能用来苛求赫歇耳，因为能量守恒定律的确立是赫歇耳去世之后的事情（在那之前只有一些关于机械能的粗糙概念）。如今连小学生都会追问的热能从何而来的问题，在赫歇耳时代并不是显而易见的。

[3] 卢瑟福最早实现的人工原子核嬗变是用 α 粒子（即氦核 ^4He ）轰击氮核 ^{14}N ，从中产生出氧的同位素 ^{17}O （同时发射出一个质子）。在那之前的1901年，卢瑟福与他的同事索迪（Frederick Soddy，1877—1956）还共同发现了原子核的自然嬗变。

[4]除这一构想外，爱丁顿还提出过一个更大胆的想法，那就是让电子和质子彼此湮灭，使它们的质量完全转变为能量。那种过程如果可能的话，将比核聚变更强大。可惜（更恰当地说是幸运）的是，那种过程是不可能发生的，因为否则的话，整个物质世界都将因电子和质子的湮灭而崩溃。而且如果恒星的能量果真来源于湮灭过程的话，将不会留下任何残骸，这与当时已经发现的作为恒星残骸而存在的白矮星相矛盾。因此爱丁顿的那个想法很快就被放弃了。不过有意思的是，几年之后，英国物理学家狄拉克（Paul Dirac, 1902—1984）出于不同的理由又重蹈了一次覆辙。

[5]需要指出的是，核聚变能够释放巨大能量的想法早在阿斯顿发现氢原子核的质量要比重元素单个核子的平均质量略大的同一年，就由法国物理学家佩林（Jean Perrin, 1870—1942）提出过。另外要指出的是，由于中子尚未被发现，在爱丁顿等人的早期设想中需要把一些电子塞进原子核里，以中和一部分质子电荷。

[6]除估计太阳核心的温度、密度等参数外，爱丁顿还分析了恒星质量与光度之间的关系，并对著名的赫罗图作出了定性解释。那些工作对于人们接受他的观点起了很大的促进作用。

[7]严格地讲，爱丁顿的假说并不是具有这些特点的唯一假说。在1896年法国物理学家贝克勒尔（Henri Becquerel, 1852—1908）发现放射性现象之后，曾有人猜测太阳的能量来自重核裂变成轻核的核裂变反应。不过人们很快就意识到太阳上没有足够多的重元素来长期提供裂变能量，因此那个猜测很快就被推翻了。

[8]值得一提的是，卢瑟福早在1921就猜测过中子的存在，用以解释原子核的稳定性（在原先的理论中，原子核是由质子和电子构成的，

它面临的一个很棘手的困难，就是禁锢在原子核中的电子的能级远高于实验观测到的核能级）。

[9] 质子-质子链中的其他几类实现方式与第一类的区别在于 ^3He 生成 ^4He 的方式各不相同。对太阳来说，那另外几类实现方式对总能量的贡献约为14%。

[10] 这是假定那一小块物质的温度始终维持在太阳的核心温度上。只有那样，它才能熔化几百千米外的钢铁。如果我们仅仅是从太阳核心挖出一小块物质，它的温度会因辐射能量而急剧降低，情形将很不相同。



绘画：张京

9 细小的粒子 巨大的谜团

在第8章中，我们介绍了作为太阳能量产生机制的爱丁顿和贝特的理论（以下简称恒星核聚变理论），它能很好地说明太阳为什么可以长时间发射如此惊人的能量，而且它是能够说明这一现象的唯一理论。从这个意义上讲，它已经经受住了初步的观测检验。但这种检验毕竟是很间接的，如果打个比方的话，有点像是通过测量一个黑箱的输出功能来检验有关它内部结构的猜测，哪怕检验合格，也未必能让所有人都信服，因为它终究不如打开黑箱直接窥视里面的结构来得确切。

科学家们想要的正是像打开黑箱直接窥视那样的确切性。

粗看起来，这胃口似乎有点贪婪，谁能打开一只被2 500亿个大气压所包围着的，具有1 570万度高温的黑箱呢？但幸运的是，在这貌似不可能的任务上，大自然却很慷慨地为我们准备了一扇窗户——一扇能让我们直接窥视太阳核心奥秘的窗户。

只不过窥视的结果却让科学家们足足困惑了30年。

我们先来说说那扇“窗户”。我们知道，普通的窗户之所以能让我们窥视东西，是因为它能让光子通过，而我们的眼睛或仪器则通过接收那些光子来获取信息。太阳核心既然是太阳系的光明源泉，发射光子自然是分内之事，而我们在地球上能见到阳光，说明我们能接收来自太阳的光子。既然一个能发射，一个能接收，粗看起来，那扇能让我们窥视太阳核心奥秘的窗户就是普通的窗户。可惜那只是错觉。我们在后文中将会看到，由太阳核心所发射的光子要经历十几万年的漫长时间才能到达

太阳表面，这个过程比唐僧取经还要艰难无数倍，而且到达太阳表面，继而抵达我们眼睛或仪器的早已不是由核聚变反应所产生的那些光子，而是不知多少代之后的“子孙”了。那样的光子虽仍能如第6章所说的那样带给了我们大量的信息，但此光子非彼光子，有关太阳核心的信息早就丧失殆尽了。

因此能让我们窥视太阳核心奥秘的窗户并不是普通的窗户，尤其光子并不是能让我们窥视太阳核心奥秘的东西。那么能使我们透过那扇窗户窥视太阳核心奥秘的究竟是什么东西呢？这还得从那些核聚变反应说起。

我们介绍过，那些核聚变反应中被称为“第一类质子-质子链”的反应提供了约85%的太阳能量，而那个核反应的第一步是“两个质子聚合成氢的同位素氘核 ^2H ”。细心的读者也许早已注意到了，那样的反应其实是不可能发生的，因为反应前后的电荷数目不同：两个质子带两个单位的正电荷，而一个氘核却只带一个单位的正电荷。那样的反应如果发生的话，现代物理学中最可靠的守恒定律之一——电荷守恒定律——就泡汤了。很明显，“两个质子聚合成氢的同位素氘核 ^2H ”那样的简略提法只关注了反应过程中的原子核部分，而忽略了反应产物中必须包含的另一个带正电的粒子。那个粒子叫做正电子，它是电子的反粒子，它的出现保证了电荷守恒。

但这还不够，因为有一样重要性不在电荷之下的东西还不守恒，那就是能量。只不过这一点不是仅凭细心就可以发现的，而必须通过实验。但早在那样的实验成为现实之前的20世纪20年代末，物理学家们就已经注意到了在与之类似——即反应产物中有一个电子或正电子——的核反应过程中能量似乎是不守恒的。[\[1\]](#)由于当时刚刚经历了量子力学革

命，物理学家们心中的“革命意识”还比较高涨，遇到那样的异常现象，有些人就开始往物理学革命的路数上去想，其中包括著名的丹麦物理学家玻尔（Niels Bohr，1885—1962），他一度猜测能量守恒定律在原子核尺度上将不再适用。他的这种“冒进”想法遭到了奥地利物理学家泡利（Wolfgang Pauli，1900—1958）的批评。1930年，泡利提出了一个不同的想法。他认为貌似不守恒的那部分能量其实是被一种尚未被观测到的细小的中性粒子带走了。那种中性粒子起初被称为“中子”，后因该名称被一种大质量的中性粒子——即我们如今所说的中子——所霸占而改称为了中微子（neutrino）。在“两个质子聚合成氢的同位素氘核 ^2H ”这一反应中，如果把中微子和正电子一起加入到反应产物中去，电荷与能量就都守恒了。

这个由泡利提出的中微子就是使我们能透过那扇窗户窥视太阳核心奥秘的家伙。

中微子这个细小的家伙为什么有那样的神通呢？原因很简单：因为它与物质的相互作用极其微弱。事实上，在主宰微观世界的三大相互作用中，它只参与其中最弱的一种，即所谓的弱相互作用。相互作用微弱所导致的一个直接后果，就是穿透能力超强。在物理实验室里，人们常常用铅来屏蔽辐射，但中微子却可以不太困难地穿越整整一光年厚的铅。正因为有如此超强的穿透能力，它才能在早年的实验中携带能量轻松逃离，给物理学家们留下一个能量不守恒的犯罪现场，甚至让玻尔那样的“神探”都栽了跟斗。^[2]也正因为有如此超强的穿透能力，它才可以取代光子成为我们窥视太阳核心奥秘的工具，因为厚达70万千米，让光子举步维艰的太阳物质对它来说几乎是完全透明的。

窗户既然开启了，那么接下来的问题就是：透过这扇窗户，我们想

窥视什么样的奥秘？在这方面，科学家们的胃口倒并不贪婪——起码到目前为止还并不贪婪，他们只想确认太阳的核心是不是真的如恒星核聚变理论所预言的那样发生着大规模的核聚变反应。确认的方法就是探测核聚变反应中必然会产生的那些中微子，即所谓的太阳中微子（solar neutrino），因为那是唯一一种能几乎不受阻碍地逃离太阳核心的粒子。

在介绍那些探测之前，让我们首先估计一下，假如恒星核聚变理论成立，太阳每秒钟会产生多少个中微子？这个估计并不困难，因为太阳上的核聚变反应虽然有很多类型，最终结果却都是将四个质子和两个电子结合成一个氦原子核，^[3]每个这样的结合过程都将发射两个中微子（感兴趣的读者不妨以第一类质子-质子链为例自己清点一下数目）。简单的比较表明，四个质子和两个电子的总质量比一个氦原子核大了26.7MeV（约合 4.27×10^{-12} 焦耳）。这表明太阳核心每产生26.7 MeV的能量，就会发射两个中微子。由于我们在第7章中已经知道太阳的光度约为 3.84×10^{29} 瓦，即每秒产生 3.84×10^{29} 焦耳的能量。由此不难算出太阳每秒发射的中微子数量约为 1.8×10^{38} 。^[4]这是一个什么概念呢？它相当于每秒——不分白天黑夜——都会有几百亿个中微子穿过你身上每平方厘米的面积！这就是爱丁顿和贝特等人的这个恒星核聚变理论所预言，而我们将要去搜集证据加以验证的“某些其他事情”。不过大家不必为这个巨大的数字而担心，虽然你一生都会“沐浴”在汹涌的中微子洪流中，却很可能一生都不会有一个中微子与你发生任何相互作用——当然，微不足道的万有引力除外。

数量如此巨大的中微子在地球附近除了来自太阳核心的核聚变反应外，再没有第二种可能了，因此只要能从实验上证实它们的存在及数量，就将是对恒星核聚变理论的判决性支持。但问题是，要想从实验上

证实中微子的存在，必须让它们在我们的探测器里留下踪迹。可中微子既然能穿透整个太阳，穿透整整一光年的铅，当然也能不留丝毫踪迹地穿透我们的探测器。它的这种超强的穿透力既是为我们提供信息的前提，同时却也是对实验技术的最大挑战。

但物理学家们还是有办法的。办法很简单，那就是“人海战术”，这其实是对付小概率事件的通用办法。我们知道，让一个人掷30次硬币要想全都掷出正面几乎是不可能的（概率只有十亿分之一），但如果让全中国十几亿人每人都掷30次硬币，那么一轮下来就会有很大的概率出现一个掷出30次正面的人，这就是“人海战术”的威力。科学家们对付中微子的办法也是如此，只不过是把“人”换成物质。中微子与物质的相互作用虽然微弱，但只要有合适的探测物质，并且数量足够多，当大量的中微子与它们擦身而过时，还是会有个别中微子留下踪迹的。

那么什么样的物质适合做探测物质呢？1946年，出生于意大利的核物理学家庞蒂科夫（Bruno Pontecorvo, 1913—1993）研究了这个问题。他提出了探测物质所应具备的一系列条件，其中包括：

- （1）不太昂贵——因为所需数量很大，太昂贵了负担不起。
- （2）与中微子反应后必须生成放射性元素，以便我们能通过放射性来判定它们的存在，进而推断反应的发生。
- （3）所生成的放射性元素必须有合适的半衰期，以便我们既有时间分离它们，又不必等太久来判定它们的存在。
- （4）除中微子外，其他因素造成同样反应的可能性必须很小，以免干扰结果。

在这样一些条件下，庞蒂科夫推荐了几种探测物质，其中最重要的一种是氯（Chlorine）的同位素 ^{37}Cl 。它与中微子反应后会生成氩（Argon）的同位素 ^{37}Ar （同时放出一个电子以平衡电荷）。 ^{37}Ar 是一种放射性元素，半衰期约为35天。

20世纪50年代，美国物理学家戴维斯（Raymond Davis, Jr., 1914—2006）率先展开了这方面的实验研究。考虑到氯在常温常压下是气体，体积过于庞大，戴维斯采用了常温常压下呈液态的四氯化碳（Carbon Tetrachloride, CCl_4 ）。他的实验室位于地下5米左右，使用的四氯化碳约为3 800升。在这样简陋的条件下，他只能得到一个有关太阳中微子数量的很宽松的上限，比理论值高出了好几个数量级。^[5]

结果虽然寒碜，毕竟算是开山之作，戴维斯还是写成一篇论文发了出去。可惜的是，论文在审稿人那里遭到了“温柔一刀”。审稿人在审稿意见中很幽默地指出：像这种缺乏精度的研究，就好比是站在山顶上，用自己的手去碰月亮，然后得出结论说月亮要比自己的手能碰得到的地方更高。审稿人的结论是：这样的研究是不应该写成论文的。

太伤自尊了。

看来必须把实验做大。从20世纪60年代中期开始，戴维斯与美国天体物理学家巴克尔（John Bahcall, 1934—2005）合作，开始在美国南达科他州（South Dakota）一座名为Homestake的金矿的矿井里建造一个巨大的中微子探测器（图9.1）。巴克尔是研究太阳模型的专家，他对太阳中微子流量的理论计算，在几乎所有太阳中微子探测实验中都是最重要的参照之一。^[6]戴维斯的这个新实验被称为Homestake实验，它的探测器位于地下1 500米深处。这种“深挖洞”的做法是太阳中微子实验中的

标准做法，目的是减少其他因素——比如宇宙线——的干扰。为了便于相互比较，人们往往用等效水深来表示中微子探测器的深度。

Homestake实验中的这个1 500米的深度用等效水深来表示大约为4 200米。在Homestake实验中，探测物质是近40万升（约600吨）的四氯乙烯（Tetrachloroethylene, C_2Cl_4 ）。



图9.1 Homestake探测器

1967年，戴维斯的实验装置基本建成。1968年，他得到了第一批观测结果，但误差很大。直到1970年，经过各种改进，他才得到了具有统计价值（即不至于被误差淹没）的结果。这结果是一个好消息和一个坏消息的混合。好消息是他的确探测到了太阳中微子，坏消息则是他探测

到的中微子流量明显小于理论预言。这个结果催生了一个新名词：太阳中微子问题（solar neutrino problem）。

Homestake实验持续进行了25年（1984年之后改由宾夕法尼亚大学主持），检测到的平均中微子流量约为理论预言的1/3。这是一个令人尴尬的结果，因为尽管检测到的流量只有理论预言的1/3，却依然是一个除恒星核聚变理论外，绝无任何其他机制可以在地球附近产生的巨大流量，从这个意义上讲，它足以作为恒星核聚变理论的铁证。但另一方面，1/3毕竟不是1，连四舍五入都入不到1。它作为定性证实虽然马马虎虎，从定量上讲，却是一个很糟糕的结果。这一点令戴维斯和巴克尔深感不安。自第一批论文发表之后，他们对实验和理论的各个方面都进行了仔细复核，试图缩小观测与理论的差距，却始终无法如愿。1989年，他们在一篇文章中写道：“对我们来说很惊讶，也许还相当失望，意识到自那些论文发表以来，尽管对细节进行了十几年的复核及持续改进，却无论在观测还是标准理论上都没什么定性的改变。”

既然自己的努力遇到了挫折，那就看看别人是怎么做的吧。物理学上的任何实验都必须能够重复，而且要尽可能由不同的实验小组、不同的实验设备、不同的实验方法、在不同的地点来重复，这样可以最有效地减少单一小组、单一设备、单一方法所可能存在的从心理因素到系统误差在内的各种不足。这种研究规范是物理学能够令人信赖的一个重要原因。

探测太阳中微子的不同方法其实早在Homestake实验开始运作之前就已经有人在研究了。1966年，俄国物理学家库兹明（Vadim Kuzmin，1937— ）提出了一种新的探测方法，利用镓（Gallium）的同位素 ^{71}Ga 作为探测物质。 ^{71}Ga 与中微子反应后会生成锗（Germanium）的同位

素 ^{71}Ge （同时放出一个电子以平衡电荷）。 ^{71}Ge 是一种放射性元素，半衰期约为11天。

利用 ^{71}Ga 来探测太阳中微子有一个Homestake实验无法比拟的优势，那就是能够探测到的中微子能量范围要宽广得多。我们在第8章中介绍过，在太阳核心所产生的能量中，第一类质子-质子链的贡献占了85%左右。毫无疑问，这是太阳核心最重要的核聚变反应，也是太阳中微子的最大来源。但这类反应所产生的中微子的最大能量只有0.423 MeV（平均能量为0.267 MeV），而Homestake实验所能探测到的中微子的最低能量——即所谓的阈能（threshold）——却是0.814 MeV，大大高于第一类质子-质子链所产生的中微子的能量。这就意味着Homestake实验对太阳中微子中数量最庞大的那部分是“视而不见”的（当然，那部分中微子在实验与理论对比时是被除去的）。考虑到检测中微子的极端困难性，这种“视而不见”无疑是极大的“浪费”。而利用 ^{71}Ga 来探测太阳中微子的最大好处就是能弥补这一缺陷，因为它所能探测到的中微子的最低能量为0.233 MeV，从而涵盖了很大一部分来自第一类质子-质子链的中微子。^[7]除这一点外，利用 ^{71}Ga 来探测太阳中微子还有一个很大的好处，那就是在理论上，人们对第一类质子-质子链的研究要比对那些所占比例小得多的其他反应彻底得多，因而能提供更可靠的理论数据。

设想是有了，将它变为现实却是一件费时费力费钱的事情，利用 ^{71}Ga 作为探测物质的太阳中微子实验直到20世纪90年代初才开始出结果。从事此类实验的有两个研究小组，一个是由苏联和美国的科学家联合进行的，称为SAGE（Soviet-American Gallium Experiment的缩写）实验，地点位于苏联高加索山区（Caucasus mountains）一条四千米深的隧道内，等效水深约为4 700米；另一个是由美、德、法、波兰、意大

利、以色列等国的科学家联合进行的，称为GALLEX（Gallium Experiment的缩写）实验，地点位于意大利阿布鲁佐大区（Abruzzo）一个等效水深约3 200米的地下实验室内（图9.2）。在实验手段上，SAGE采用的实验物质是30吨液态镓（1991年增加到57吨），GALLEX采用的则是101吨三氯化镓（gallium trichloride）溶液。

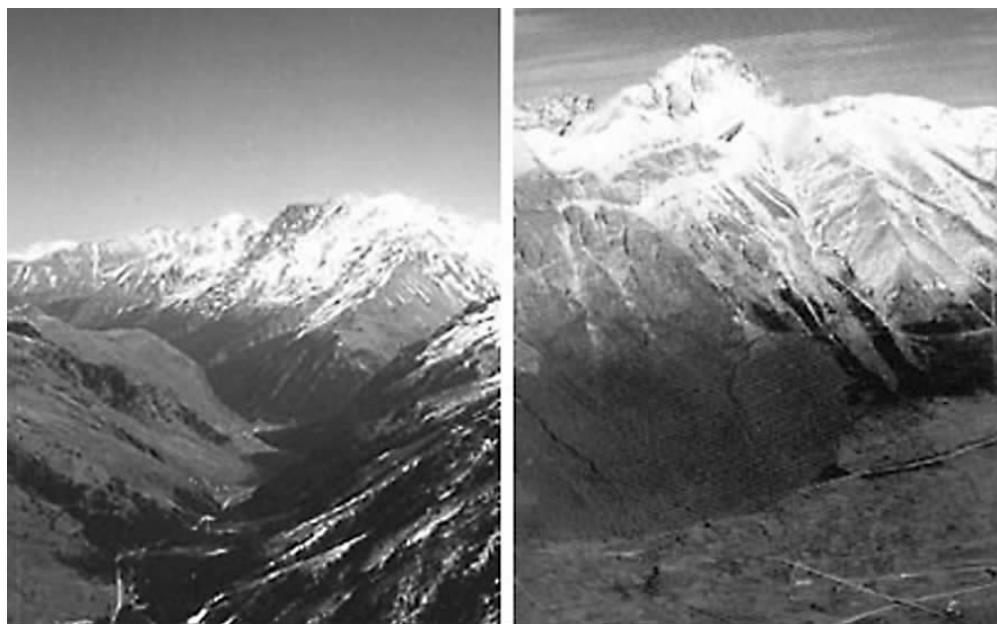


图9.2 SAGE（左）和GALLEX（右）实验所在的山区

那么这两组新实验的结果如何呢？它们是“A. 找到了全部的太阳中微子”呢？还是“B. 证实了Homestake实验那缺斤短两的结果”？答案是：“C. 以上都不对”。这两组新实验探测到的太阳中微子流量既不是 $1/3$ ，也不是1，而大约是理论流量的60%。[\[8\]](#)

有点盲人摸象的意味了，你摸到一个脑袋，我摸到一条大腿。真相到底如何呢？还得再摸摸看。

另一组“摸象”的“盲人”是日本人，他们的实验结果其实比SAGE和GALLEX出得更早，他们的实验地点在日本神冈町（Kamioka）的一座

等效水深约2 700米的地下矿井内，称为神冈观测站（Kamioka Observatory），原先从事的是一度很热门的质子衰变研究。自1985年开始，该观测站经过扩建后改称为神冈核子衰变实验Ⅱ期（Kamioka NDE-Ⅱ）。自1987年起，该实验室的探测器——神冈探测器——开始进行太阳中微子探测。

神冈探测器探测太阳中微子的方式与Homestake、SAGE、GALLEX等实验都不相同，它核心部位的探测物质是2 142吨高度纯净的水，所利用的反应过程则是中微子与电子的碰撞。这种碰撞当然是小概率事件（因此要用几千吨水），但如果发生了，就有可能使电子获得能量，如果入射中微子的能量很高，电子所获得的能量也会很多，运动速度甚至可以超过光速——别紧张，只是超过水中的光速，没有破坏相对论。人们在电磁学研究中早就知道，当电子在水中的运动速度超过水中的光速时，会发射一种特殊的辐射，叫做切连科夫辐射（Cherenkov radiation）。通过观测这种辐射，物理学家们就可以确定反应的发生。

神冈探测器的这种与众不同的探测原理使它具有一种Homestake、SAGE、GALLEX等实验都不具有的优势，那就是可以确定中微子与电子发生反应的时间、位置、入射方向、入射能量等细节。这些细节对于统计太阳中微子的数量来说虽不是必须的，但对于深入探索太阳核心的其他奥秘却有很大价值。^[9]不过有得就有失，神冈探测器也有一个很大的问题，那就是阈能特别高。为了使碰撞后的电子能够“超光速”，并且发射足够强劲的切连科夫辐射，入射中微子的最低能量要达到7.2 MeV，只有“极少数极少数”的太阳中微子能具有如此高的能量。

从1987年到1990年，神冈探测器在积累了1 040天的数据后得到了一个结果：它探测到的太阳中微子流量约为理论流量的46%。1995年，

在积累了2 079天的数据基础上，上述结果被修正为55%。1996年，神冈观测台耗资一亿美元建造了更大的探测器，称为超级神冈（Super-KamiokaNDE）探测器（图9.3），它的探测物质增加到了50 000吨高度纯净的水，它对切连科夫辐射的探测灵敏度达到了可以探测到月亮上一支烛光的惊人程度，而它的阈能则降低到了5.5 MeV。一句话，它在各方面都有了长足的改进。自1998年起，超级神冈探测器开始发布探测结果，它所探测到的太阳中微子流量约为理论流量的47%，与其他各组实验都不相同。

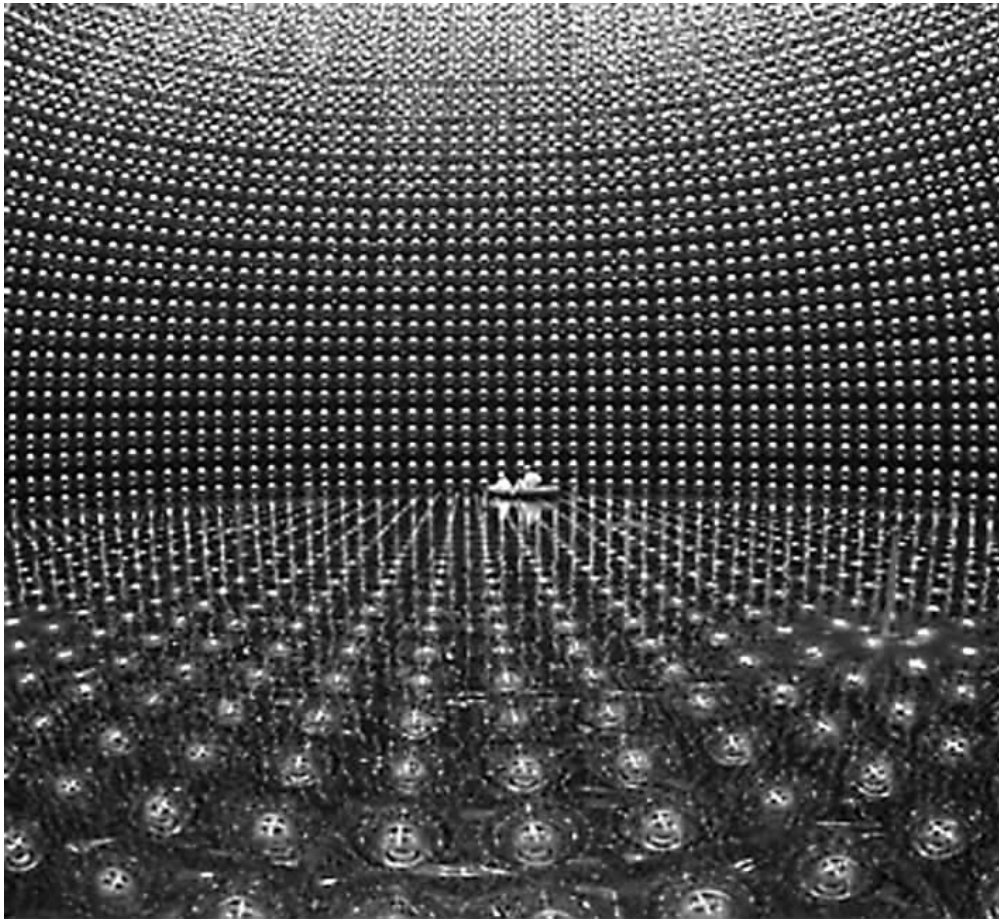


图9.3 超级神探测器

杯具了！

20世纪就在这样忙忙碌碌的太阳中微子研究中走到了尾声。虽然在这类研究中，无论实验还是理论都存在不小的误差，但有一点已经可以确定，那就是探测到的太阳中微子流量明显小于理论预言，而且两者在不同的能量范围内有不同的差距。中微子这个细小的粒子，给即将到来的新世纪留下了一个巨大的谜团。究竟是太阳原本就没有发射那么多中微子呢？还是有些太阳中微子失踪了？如果是失踪，那它们究竟跑到哪里去了呢？

[1]这里所说的“反应产物中有一个电子或正电子”的核反应过程是指当时人们正在积极研究的重原子核的 β 衰变，那种反应过程中的能量异常现象其实更早些时候就被注意到过，但直到20世纪20年代末，随着其他可能性被陆续排除掉，才真正威胁到了能量守恒定律。另外要说明的是，那些实验所导致的只是中微子的提出，包括正电子在内的反粒子概念的提出则是出于其他原因，时间上也更晚一些。

[2]中微子直到1956年才由美国物理学家考恩（Clyde Cowan, 1919—1974）、莱因斯（Frederick Reines, 1918—1998）等人所发现。莱因斯因此而获得了1995年的诺贝尔物理学奖（考恩不幸已经去世）。

[3]确切地讲，这只是现阶段的主要反应模式。在后文中我们将会看到，当太阳步入“晚年”后情况将会有所变化。另外要说明的是，参与反应的两个电子的参与方式不止有一种：主要的方式是与反应产物中的正电子湮灭；稀有的方式是直接参与反应，那样在反应产物中就不会有正电子（因为电荷已经平衡了）。

[4]这个估算也请读者自己去做。不过要提醒的是，这个估算是有一定系统误差的，感兴趣的读者请思考一下，什么因素会造成系统误差？它

会使估计出的中微子数量偏大还是偏小？

[5]他所得到的这一上限为40 000 SNU。SNU是所谓的太阳中微子单位（Solar Neutrino Unit），它描述的是中微子探测器探测太阳中微子的能力。一个SNU表示每秒每 10^{36} 个探测器原子可以俘获一个太阳中微子。

[6]有读者也许会问：我们前面不是已经很轻松地计算出了太阳中微子流量吗？何必还要什么专家呢？答案是：我们计算的只是太阳中微子的总量，对实验探测来说重要的不是总量，而是中微子数目随能量的分布，因为各个探测器所能探测到的中微子能量范围各不相同，在不同能量上的探测灵敏度也各不相同。中微子数目随能量的分布与太阳内部各种核聚变反应的相对比重等各种复杂因素有着密切关系，是非常困难的理论问题。

[7]这一点从两者所能俘获的太阳中微子流量的大小可以看出。Homestake实验所能俘获的理论流量约为8 SNU，而以 ^{71}Ga 为探测物质的探测器所能俘获的理论流量高达130 SNU左右。

[8]由于对太阳中微子流量的理论计算存在几种彼此接近，但互不相同的结果，因此视所用理论结果的不同，不同文献给出的实验与理论的比值存在一定差异。

[9]其他几种探测器之所以做不到这一点，是因为它们反应产物的半衰期长达十几或几十天，等到人们能通过反应产物的衰变来确定反应的发生时，有关反应的细节信息早已不复存在了。



绘画：张京

10 标准太阳模型vs粒子物理标准模型

由太阳中微子带来的这些问题被称为太阳中微子问题（solar neutrino problem），有时也被称为太阳中微子之谜，或太阳中微子失踪之谜。不过严格来说，这最后一个名称是需要论证的，这就好比在现实世界中要把一个案件定性为失踪案，是需要经过排查的。在本章中，我们就从那些排查说起。

我们要排查的第一种可能性是观测出问题的可能性。这种排查之所以必要，是因为中微子是一种相互作用极其微弱，从而在观测上极易“漏网”的粒子。这“漏网”达到一定程度，就有可能无中生有地产生出太阳中微子问题来，就像早年的核物理实验因为无法探测到中微子，而无中生有地闹出了能量不守恒的伪问题来一样。

那么，太阳中微子问题是否也有可能是那样的伪问题呢？答案是否定的。理由有两条：一是太阳中微子问题乃是几组独立实验的共同结果，这种结果的可靠性要比单一实验大得多。二是GALLEX和SAGE这两组实验都用流量已知的人工中微子源对探测器进行过校正，从而进一步确保了它们的可靠性。因此我们有充分的理由相信，太阳中微子问题并不是观测错误导致的伪问题。

既然不是伪问题，那就是真问题了；而既然不是观测问题，那就是理论问题了。因此，接下来要排查的就是哪部分理论出了问题。我们知道，对太阳中微子流量的理论预言来自所谓的太阳模型，这是一种在学术界有着广泛共识的模型，也称为标准太阳模型（standard solar model）。这一模型以太阳的大小、光度、表面温度等可观测数据为约

束，对太阳内部物态的分布、压强的平衡、能量的产生与传输、辐射的吸收与发射等主要因素及相互关系给出了定量描述。它决定着太阳核心产生中微子的机制及数量。但标准太阳模型并不是故事的全部，因为太阳中微子被产生之后还必须经历一个传播环节：它必须穿越几十万千米的太阳物质以及日地之间将近一亿五千万千米的距离，才能进入我们的探测器。而描述这一传播环节的是一个有关粒子物理的模型，那也是一种在学术界有着广泛共识的模型，叫做粒子物理标准模型（standard model of particle physics）。这一模型对目前已知的所有基本粒子及其相互作用作了相当精密的描述。因此，与太阳中微子问题有关的理论其实有两个部分，它们拥有一个共同的名字叫做“标准模型”。而所谓“哪部分理论出了问题”，归根到底是两个标准模型之争，即标准太阳模型粒子物理标准模型。

这两个标准模型哪个更可靠呢？在谜底揭晓之前如果让物理学家们押宝，我想绝大多数物理学家^[1]会把宝压在粒子物理标准模型上。因为这个模型自20世纪60年代后期建立以来，已经得到了无数的检验，其中包括所预言的新粒子及其参数得到证实那样堪称经典的检验。粒子物理标准模型所描述的虽然是肉眼无法看见的微观世界，但对物理学家们来说，实验室里数不胜数的粒子反应图片无时无刻不在述说它的实在性。与之相比，标准太阳模型所受到的检验却少得可怜，其中被寄予厚望的太阳中微子实验偏偏又得出了令人尴尬的结果。而且与粒子物理标准模型所描述的相对纯粹的微观世界不同，标准太阳模型涉及的是一个巨型天体的内部世界，只要想想我们脚底下的地球尚且如此复杂，我们对标准太阳模型的信心就很难不打上几分折扣。因此直到20世纪90年代，曾因研究大统一理论而闻名的美国物理学家乔治（Howard Georgi, 1947—）还在一篇论文中宣称太阳中微子问题与粒子物理无关，他并且很体贴地表示：天体物理学家们能够把太阳中微子的数目计算到只差两到

三倍的程度，就已经很了不起了。言下之意，实验与理论的这点出入是不足为奇的，可以由标准太阳模型负全责，别来烦粒子物理。另一位知名的美国物理学家德雷尔（Sidney Drell，1926— ）也表示，粒子物理标准模型已经辉煌到了难以被放弃的程度。

既然“民意”如此，那我们就先考虑标准太阳模型出问题的可能性吧。我们在第8章中提到过，太阳核心核聚变反应的剧烈程度与太阳的核心温度有着极为敏感的依赖性。由于太阳中微子来自于核聚变反应，它的流量当然也与后者的剧烈程度，从而与太阳的核心温度有着极为敏感的依赖性。因此，标准太阳模型出问题的最大可能性，就是它所预言的太阳核心温度出了错。如果太阳核心的实际温度比标准太阳模型所预言的低，那么太阳核心核聚变反应的剧烈程度，以及它所产生的太阳中微子的流量就会大幅降低。计算表明，若太阳核心温度与核聚变反应之间的敏感依赖性的福，太阳的核心温度只要调低几个百分点，就足以使太阳中微子的流量减少几十个百分点，从而与观测结果定性相符。

初看起来，这是一种很大的可能性，因为在描述像太阳核心那样远离经验的环境时，出现几个百分点的温度误差不仅是完全可能的，甚至可以说是无可避免的。但细想一下却又不然，理由很简单：将太阳的核心温度调低几个百分点虽然能定性地调和中微子流量的理论与观测之间的差距，同时却也会导致太阳光度的大幅降低——因为核聚变反应的剧烈程度大幅降低了，而这是与观测完全相悖的。

不过仅凭这一点还难不倒物理学家们，因为有一些方案可以在调低太阳核心温度的同时维持太阳光度不变。这其中比较拙劣的方案是利用我们在第9章中提到，并将在后文中进一步介绍的“由太阳核心所发射的光子要经历十几万年的漫长时间才能到达太阳表面”这一特点。利用这一特点，有人提出了一个很离奇的假设，那就是太阳的核心温度是在

不久之前才突然变低的，由此导致的效果尚未传到太阳表面，从而尚未影响太阳光度。这个假设的拙劣之处是显而易见的，因为它是为解决太阳中微子问题而特意炮制的，而且它甚至比太阳中微子问题更难理解。这种用一个比原始问题更难理解的假设来解决问题的做法几乎是与科学的宗旨背道而驰的。连这样的“天方夜谭”都被提出来了，可见人们是多么希望把问题归咎于标准太阳模型。

不过，修改标准太阳模型的方案也并不都是拙劣的。相对高明的方案是在调低太阳核心温度的同时扩大核反应区域的范围，以抵消核反应剧烈程度降低造成的影响，从而达到维持光度不变的目的。但不幸的是，我们在第9章中已经介绍过，太阳中微子的总流量是与太阳光度直接对应的。一旦维持太阳光度不变，也就维持了太阳中微子的总流量不变。因此这种方案至多能解释某些特定能量范围——比如高能区——内的太阳中微子问题（因为核心温度的调低会影响各种核聚变反应的相对比例，从而影响不同能量中微子的相对比例——尤其是减少高能中微子的数量）。但随着实验结果所涵盖的能量范围越来越宽，太阳中微子问题早已不仅仅是特定能量范围内的问题，而变成了总流量上的问题，这就超出了维持太阳光度不变的那些方案的解释范围。^[2]

但对修改标准太阳模型的方案构成更沉重打击的，则是我们在后文中将要介绍的所谓日震学研究，那种研究证实了标准太阳模型的许多细节，从而为修改标准太阳模型的努力设置了相当苛刻的限制。事实上，标准太阳模型看似“五大三粗”，只用寥寥几个方程式来描述整个太阳的基本特征，但它对物理原理的运用却是相当缜密的，甚至达到了“牵一发而动全身”的精密程度。比如要想调低太阳的核心温度，就必须同时调节太阳内部重元素比例之类的参数，那些参数有不少是可以用日震学手段进行检验的，而检验的结果几乎无一例外地支持了标准太阳模型。

除日震学研究外，对标准太阳模型的支持还来自另外一个领域，那个领域不是别的，正是把怀疑目光引向标准太阳模型的太阳中微子研究本身。我们说过，标准太阳模型出问题的最大可能性，就是它所预言的太阳核心温度出了错。之所以这么说，首要原因当然是太阳核心温度与中微子流量有着极为敏感的依赖性，从而在表观上具备解决太阳中微子问题的潜力，但另一方面也是因为我们无法直接测定太阳核心温度，从而无法进行实验裁决。

实验鞭长莫及的地方，往往就是理论遍地开花的地方。

但细心的读者也许还记得，我们在第9章中曾经把中微子表彰为“取代光子成为我们窥视太阳核心奥秘的工具”。既然给予了这么高的荣誉，我们就要问：在它让我们窥视到的“太阳核心奥秘”中，是否包括了太阳的核心温度呢？很幸运，答案是肯定的，就像光谱可以告诉我们发光体的温度一样，太阳中微子的能谱也可以带给我们有关太阳核心温度的信息。这一答案，加上像神冈探测器那样能够测定中微子能量的探测器的问世，使我们可以通过太阳中微子的能谱来推算太阳的核心温度。而推算的结果与标准太阳模型的符合程度远高于修改方案。

自此，修改太阳模型的努力算是撞上了南墙，虽不能说从此无人问津，但希望已变得很渺茫了。而太阳中微子问题也确实可以定性为失踪案了，因为标准太阳模型既然没有问题，就说明太阳中微子的流量确实要比实验探测到的大，一部分太阳中微子确实失踪了。

现在留给我们的就只剩一条路了：修改粒子物理标准模型。粒子物理标准模型再是“辉煌到了难以被放弃的程度”，到了这个地步，我们也只得在它头上动土了。不过在动土之前，我们先要对失踪者——中微子——的家世做一个彻查。据我们目前掌握的情况，中微子家族共有三兄

弟，个个都是偷运能量的好手。我们前面所说的中微子只是最早落网的那个，它的全名叫做电子中微子（electron neutrino）。除电子中微子外，人们在1962年和2000年又先后发现了另外两种中微子，分别叫做 μ 子中微子（muon neutrino）和 τ 子中微子（tau neutrino）。^[3]这种三兄弟结构为从修改粒子物理标准模型角度解决太阳中微子问题提供了一条重要思路，叫做中微子振荡（neutrino oscillation）。

什么是中微子振荡呢？简单地讲就是中微子三兄弟之间的相互转变——比如电子中微子变为 μ 子中微子， μ 子中微子变为 τ 子中微子等。这种转变可以循环往复，因而称之为振荡。中微子振荡之所以有可能解决太阳中微子问题，是因为太阳核心所产生的全都是电子中微子，而我们在第9章中介绍过的那些中微子探测器所探测的也主要是电子中微子。^[4]因此假如来自太阳核心的电子中微子在飞往地球的途中有一部分转变成了 μ 子中微子或 τ 子中微子，就会逃过探测器的检验，而造成中微子失踪的假象。这就好比一位用几个马甲（笔名）轮流发帖的网民，从IP地址上显示他一共发了一百个帖子，但你若只搜索其中一个马甲，就只能搜到一部分帖子。

有关中微子振荡的想法有着不短的历史。事实上，早在中微子的兄弟被发现之前，也早在太阳中微子问题出现之前，就有人提出了中微子振荡的想法，此人就是我们在第9章中提到过的那位出生于意大利的核物理学家庞蒂科夫。1958年，庞蒂科夫提出了中微子和反中微子相互振荡的可能性。^[5]在 μ 子中微子被发现之后，1967年，他与时俱进地修改了自己的猜测，提出电子中微子与 μ 子中微子也有可能发生振荡。从某种意义上讲，他的这些猜测可以说是预言了太阳中微子问题的出现，因为中微子振荡必然会导致只对电子中微子敏感的中微子探测器无法探测到全部中微子。当戴维斯的Homestake实验让太阳中微子问题粉墨登场

后，庞蒂科夫又与俄国物理学家格里波夫（Vladimir Gribov，1930—1997）一起，在第一时间重申了自己的猜测。

当然，通过修改粒子物理标准模型来解决太阳中微子问题的方案并非只有中微子振荡这一种，只不过其他方案都很短命，比如有一种方案叫做中微子衰变。它认为中微子之所以失踪，不是因为披上了其他马甲，而是半路夭折了。在粒子物理中，粒子的夭折有一个很文雅的名字叫衰变。^[6]可惜的是，这种方案在1987年遭到了判决性的否认。那一年，包括神冈探测器在内的几组中微子探测器观测到了一批来自大麦哲伦星云（Large Magellanic Cloud）中的一次超新星爆发——著名的超新星1987A——的反中微子。那些反中微子经过了近17万年的长途跋涉才抵达地球。由于粒子与反粒子有着相同的寿命，如果中微子在从太阳到地球的区区8分20秒时间里就会夭折掉一部分，那我们是绝不可能观测到那么多来自超新星1987A的反中微子的。因此，中微子即便会夭折，也绝不可能那么快地夭折。既然中微子不能那么快地夭折，那就只好让方案本身夭折了。

真正有希望的还是中微子振荡。

那么，中微子振荡需要从哪里入手修改粒子物理标准模型呢？是从添加中微子的质量^[7]入手。在粒子物理标准模型中，所有中微子都是无质量的。在理论上可以证明，如果所有中微子都是无质量的，中微子振荡就不可能发生。^[8]因此，要想有中微子振荡，就必须从添加中微子质量入手，修改粒子物理标准模型。

接下来我们想知道的是：在中微子振荡中，中微子相互转变的概率满足什么样的规律？因为只有知道了这一点，才能与观测相比较。简单的理论分析表明，中微子相互转变的概率与中微子的质量、能量、相互

间的混合，以及飞行距离等诸多因素都有关联。^[9]这些关联不仅为拟合观测数据提供了许多可调节的参数，而且也定性地解释了不同能量的中微子有不同的失踪比例这一现象。就凭这两点，它已经比我们前面提到过的那些方案更有希望了。

可惜那只是表面现象。

后来当人们更细致地研究了中微子振荡之后，却发现了一些问题。比如计算表明，要想用中微子振荡来解释观测结果，需要假定中微子的某些参数与像日地距离那样与中微子风马牛不相及的参数之间满足一定的巧合关系。这种有赖于巧合的解释是物理学家们素来不喜欢的，因此是一个不好的兆头。如果说这还只是口味问题，那另外几个问题可就都是“你死我活”的级别了。比如计算表明，由中微子振荡所导致的中微子失踪率应该呈现明显的季节变化（这是因为中微子相互转变的概率与飞行距离有关，而在不同季节里，日地距离是不同的），但观测却并未发现与理论预期相一致的变化。此外，中微子相互转变的概率与能量有关这一特点虽然定性地符合不同能量的中微子有不同失踪比例这一特点，但在定量上，观测所显示的失踪率与能量的相关性——尤其在高能区——却明显小于理论的预期。

这样一来，情况就又有些不妙了。如果连中微子振荡也行不通，那岂不是连最后一条路也要变成绝路？幸运的是，在中微子振荡中还有一个重要因素起到了“挽狂澜于既倒”的作用，那就是物质对中微子振荡的影响。虽然没有明说，我们前面提到的中微子振荡其实都是指真空中的振荡，但事实上，太阳中微子一出世就得穿越厚达几十万千米的太阳物质，这会对它们产生一定的影响。有读者也许会问：我们在第9章中不是说过，太阳物质对中微子来说几乎是透明的吗？是的，我们曾经说

过。但透明不等于完全没有影响，玻璃对光来说也是透明的，却可以导致各种光学效应。太阳物质对于中微子来说也是如此，虽然透明，但会有一定的影响——尤其是对高能中微子会有一定的影响。而且更重要的是，太阳物质对不同类型的中微子有着不同的影响，对电子中微子的影响要比对 μ 子中微子和 τ 子中微子大，由此导致的后果是对中微子振荡——尤其是对高能中微子的振荡——产生影响。这种影响早在1978年，就由美国物理学家沃芬斯坦（Lincoln Wolfenstein, 1923— ）研究过。1985年，苏联物理学家米克耶夫（Stanislav Mikheyev）和斯米诺夫（Alexei Smirnov, 1951- ）推进了这一研究，并得到了一些重要结果。这种物质对中微子振荡的影响因此而被冠以他们三人的姓氏首字母，称为MSW效应（MSW effect）。后来的定量计算表明，考虑了MSW效应后的中微子振荡与所有的太阳中微子观测都相容，并且是唯一一种与所有观测都相容的解释。

不过，中微子振荡作为太阳中微子问题的正解，它的真正确立则是建立在更直接的实验裁决之上的。这种裁决首先来自日本的超级神冈探测器。1998年，超级神冈探测器通过检测宇宙线在地球大气层中产生的 μ 子中微子与电子中微子的相对数目，首次直接证实了中微子振荡的存在。^[10]在太阳中微子研究中，神冈系列探测器可谓居功至伟，占据了不止一项第一。除首次直接证实中微子振荡外，我们在第9章中提到的它能测定中微子入射方向这一特点（它对中微子振荡的证实也有赖于这一特点），还使它成为了第一个直接证实太阳中微子来自太阳方向的探测器（此前的其他探测器都只能从数量上间接推断中微子来自太阳，因为其他来源都不可能如此巨大）。而在前面提到的对来自超新星1987A的反中微子的探测中，它立的也是头功。

超级神冈探测器虽然证实了中微子振荡的存在，但对于解决太阳中

微子问题来说却还不够一锤定音，因为它在证实中微子振荡时所观测的并不是太阳中微子，而是能量远高于太阳中微子的由宇宙线产生的中微子。

那么，对太阳中微子问题最一锤定音的裁决来自哪里呢？来自加拿大的萨德伯利中微子观测台（Sudbury Neutrino Observatory, SNO），如图10.1所示。这座位于等效水深6 000米处的观测台所用的观测物质是1 000吨重水。这“重水”与超级神冈探测器所用的“水”虽只有一字之差，对于太阳中微子探测来说却有天壤之别。我们知道，水分子是由一个氧原子与两个氢原子组成的，而重水“重”在哪里呢？就“重”在那两个氢原子被换成了氘原子上。氘原子的原子核是由一个质子和一个中子组成的。当中微子与那样的原子核相遇时，有一定的概率会发生所谓的中性流反应（neutral current reaction），其结果是“棒打鸳鸯”——将氘核拆散成单独的质子和中子。与其他探测器所依赖的反应不同，这种中性流反应对三种中微子是完全一视同仁的。因此无论中微子披的是什么马甲，都能被检测到。更妙的是，除中性流反应外，中微子与重水还可以发生其他类型的反应，那些反应则与其他探测器一样，只对电子中微子敏感。^[11]因此，萨德伯利中微子观测台可以在测定中微子总流量的同时确定电子中微子的比例，从而既验证标准太阳模型的预言，又检验中微子振荡。

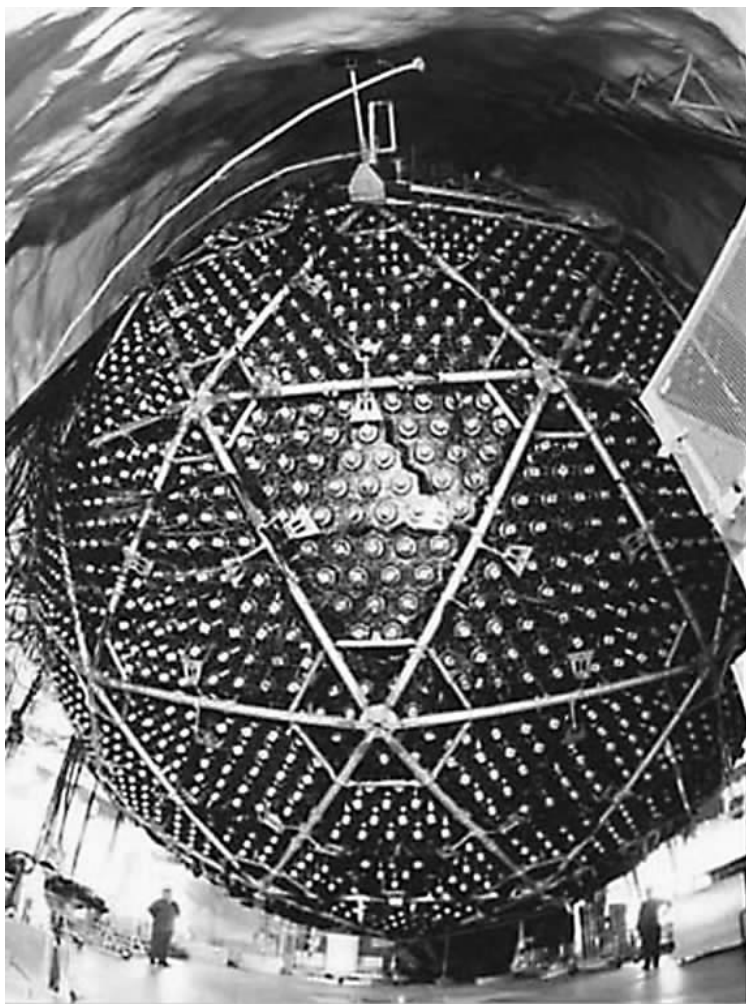


图10.1 SNO探测器

2001年，萨德伯利中微子观测台发布了观测结果，非常漂亮地显示出中微子的总流量在实验精度范围内与标准太阳模型相符，而电子中微子所占的比例，则与中微子振荡所预期的一致。

这样，太阳中微子之谜就被解开了，谜底就是中微子振荡。在标准太阳模型与粒子物理标准模型的对决中，居然是标准太阳模型取得了胜利，这让很多人跌破了眼镜。此后，人们又通过大量的其他实验（包括使用来自加速器及核反应堆的中微子）进一步证实了中微子振荡。经过如此曲折的努力，我们终于完成了第8章末尾所说的为恒星核聚变理论

收集进一步证据的工作，使这一理论得到了牢不可破的确立。

2002年，那位半个世纪前被审稿人揶揄过的“站在山顶上，用自己的手去碰月亮”的太阳中微子探测先驱戴维斯以88岁的高龄，与超级神冈探测器的“头领”，日本物理学家小柴昌俊（Masatoshi Koshiya, 1926— ）一同荣获了诺贝尔物理学奖。

2004年，戴维斯当年的合作者、太阳模型专家巴克尔撰写了一篇评述太阳中微子问题的文章。在文章的末尾，他这样写道：

当我回顾过去四十年在太阳中微子研究领域所取得的成就时，我感到了惊讶。由数以千计的物理学家、化学家、天文学家和工程师组成的国际团队用他们的合作，展示了通过统计地下矿井里一个盛满纯净液体的游泳池里的放射性原子的数目，就能告诉我们有关太阳核心的重要事实，以及被称为中微子的奇异基本粒子的性质。若非亲身经历了太阳中微子的传奇，这对我来说将会是难以置信的。

一段传奇虽已落幕，但围绕它的谜底却有许多新的问题有待探索，比如中微子的质量从何而来？中微子的质量与电子等粒子的质量是否是同一类型的？等等。一个小小的中微子尚且有如此多的奥秘，更何况太阳呢？因此大家不要离开，我们的太阳故事还将继续。

[1] 研究太阳中微子问题涉及理论、实验及观测等诸多方面，参与者来自许许多多不同领域，但为行文简洁起见，本节将把参与者统称为“物理学家”，这里预先向其他各“家”致以歉意。

[2] 在这里顺便提一下，在调低太阳核心温度的同时除了要维持太

阳的光度外，还必须维持它的核心压强，因为那是抵抗引力坍塌的资本。为了维持太阳的核心压强，物理学家们也引进了一些假设，比如假设太阳核心具有速度惊人的自转，从而能用离心力来抵御一部分引力；或假设太阳核心存在强度惊人的磁场，从而能用磁场压来抵御一部分引力。这些假设各有各的问题，比如核心的高速旋转会使太阳的整体形状及引力四极矩发生变化，与观测矛盾；而太阳核心的强磁场即便存在，也会很快衰减。

[3]这三种中微子分别用电子、 μ 子及 τ 子这三种带电轻子来命名不是偶然的，它们之间的——对应是粒子物理标准模型的基本特征之一。

[4]确切地讲，在太阳中微子的能量范围内，Homestake、SAGE及GALLEX探测器都只能探测电子中微子，神冈及超级神冈探测器则有一定的能力探测 μ 子中微子和 τ 子中微子，但敏感度只有探测电子中微子的15%左右。

[5]读者也许会觉得奇怪，在太阳中微子问题尚未出现，实验上又还没发现第二种中微子的年代，庞蒂科夫同学为什么要提出中微子振荡呢？那是因为不久之前，人们刚刚发现了中性K介子系统中的粒子振荡现象。由于中微子也是中性的，因而庞特考沃猜测它也具有类似性质。

[6]有读者也许会问：中微子已经很“微”了，它还能衰变成什么呢？一般认为，那可能会是标准模型中不存在的所谓右旋中微子或左旋反中微子。那种粒子比中微子更不像话，连弱相互作用都不参与，因而被称为惰性中微子（sterile neutrino）。不过，惰性中微子若是存在，其质量有可能相当大，在有些模型中甚至可以达到大统一能标（ 10^{15}GeV ）的量级，比质子还重 10^{15} 倍。因此，以它为衰变产物的中微

子衰变其实在理论上就不太可能。

[7] 若无特殊说明，当我们提到微观粒子的质量时，指的都是静质量。

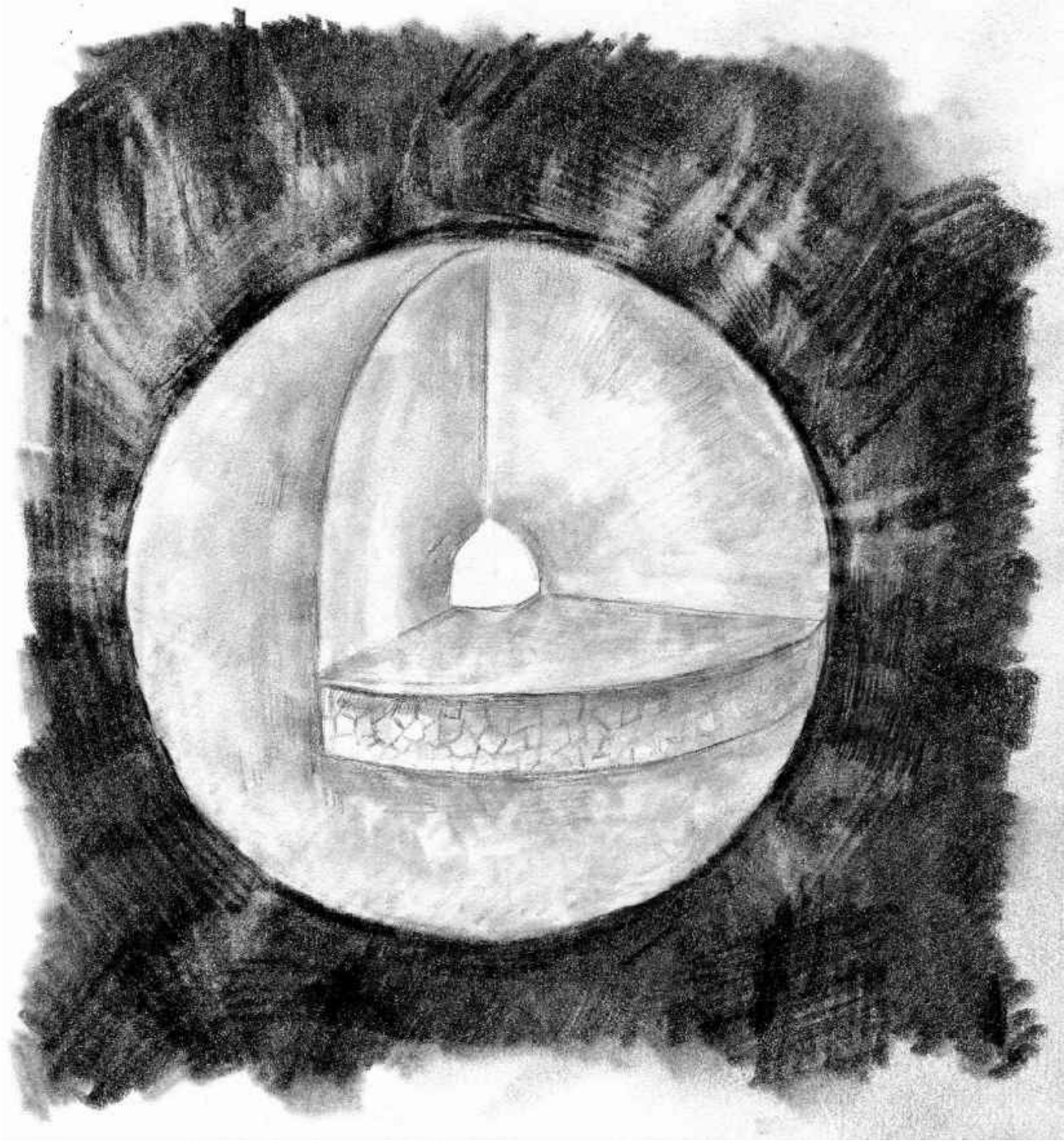
[8] 这是因为，我们通常所说的电子中微子、 μ 子中微子和 τ 子中微子都是依据它们参与弱相互作用的方式来区分的，用物理学家们的术语来说，是所谓的弱本征态。在中微子全都无质量的情形下，那就是区分中微子的唯一方式。但假如中微子有质量，我们就可以引进另一种区分方法，即按照质量来区分。那样区分出来的中微子状态被称为质量本征态。一般来说，弱本征态与质量本征态是彼此混合但互不相同的，通俗地讲就是“你中有我，我中有你”，这正是中微子振荡的温床。因此中微子振荡只有在中微子有质量（并且质量彼此不同）时才会发生。

[9] 确切地说，是与中微子质量本征值的平方差、中微子的能量、中微子弱本征态与质量本征态之间的混合矩阵（mixing matrix），以及中微子的飞行距离等因素有关。

[10] 超级神冈探测器所检测的是宇宙线在地球大气层中产生的 μ 子中微子与电子中微子的相对数目随中微子入射方向的变化。由于入射方向不同的中微子从产生到被检测之间的飞行距离各不相同（两个极端情形是：自上而下入射的中微子的飞行距离约为大气层的厚度，即十几千米；自下而上的中微子的飞行距离则为整个地球的直径，即约13 000千米），如果存在中微子振荡，它们应该有不同转变概率。通过研究这种转变概率与飞行距离的关联，物理学家们不仅证实了中微子的振荡，而且对某些中微子参数进行了估计。

[11] 具体地讲，除中性流反应外，中微子与重水之间还可以发生另

外两类反应：一类是所谓的带电流反应（charged current reaction），它使氘核分裂为两个质子（同时放出一个电子以平衡电荷），在太阳中微子的能量范围内，这种反应只能由电子中微子诱发，从而只能检测电子中微子。另一类则是与神冈及超级神冈探测器所用相同的反应，即中微子与电子的碰撞，它对电子中微子最敏感，但同时也有15%左右的相对敏感度检测到 μ 子中微子和 τ 子中微子。



绘画：张京

11 光子大逃亡

在前面几章中，我们介绍了发生在太阳核心区里的太阳能量产生机制。这一机制不仅在理论上可行，而且经过对太阳中微子的细心探测，以及对太阳中微子问题的艰辛求解，在观测上也得到了很漂亮的确立。从某种意义上讲，隐藏在太阳最深处的那个最远离经验的“恐怖核心”，反而可以说是成为了整个太阳结构中被我们了解得最可靠的部分。

如果说迄今为止我们的太阳故事所展现的大都是太阳研究中的坚实大地——那些被观测或实验牢牢确立了的事实或理论——的话，那么从本章开始，我们将会更多地去欣赏太阳研究中的绚烂天空——那些尚在云端里的谜团。我们将会看到，那样的谜团简直是层出不穷，而且在绝大多数谜团面前，我们再也没有像解决太阳能量产生机制或太阳中微子问题那样的好运气了，因为那些谜团中的绝大多数直到今天依然是未解之谜。当然，这本身未尝不是一种好运气，尤其是对于正在从事或有志于从事太阳研究的人来说更是如此，因为生在一个有许多未解之谜可以探索的时代里，要远比生活在一个只能在“小数点后第六位数字”上做文章的时代幸运得多。

好了，现在继续我们的太阳故事。

读者们想必还记得，在我们的注意力被喧宾夺主的中微子吸引走之前，我们是在谈论与阳光有关的话题，比如太阳的光谱、太阳的光度、太阳为什么会发光等。现在我们要把注意力转回到阳光上来——毕竟，那是太阳之所以成为太阳的最重要特征，也是它对我们最至关重要的东西。我们已经知道，阳光的巨大能量来源于发生在太阳核心区的核聚变

反应。那个核心区有多大呢？细致的研究表明，它从“ground zero”算起，向外延伸到大约四分之一太阳半径——17万千米——处。从体积上讲，这一区域在整个太阳中所占比例不到2%，但由于密度超高，占太阳质量的比例却达到一半左右。在这一区域的边缘，太阳物质的密度从最中心的160克/厘米³左右降到了约20克/厘米³（略高于黄金的密度），温度则降到了800万度左右。那样的环境虽然仍很恐怖，对于氢核聚变成氦核的核聚变反应来说却已低得有些勉强了，在那以外，核聚变反应就基本绝迹了。[\[1\]](#)

但这个边缘对于太阳能量的主要载体——诞生于恐怖核心的无数光子来说，却是一段新的征程——一条逃亡之路——的开始，它们将用十几万年的漫长时间，穿越厚达几十万千米的太阳物质，把巨大的能量带出去，让“红星”照耀四方。

与中微子的“挥一挥衣袖，不带走一片云彩”相比，光子的逃亡之路可就艰辛多了。原因很简单，因为光子会参与电磁相互作用——即与带电粒子发生相互作用。而更糟糕的是，在太阳的恶劣环境下，物质粒子间存在着极为猛烈的相互碰撞，其结果是“鸡飞蛋打”——几乎每个原子都会被碰掉一些电子。那些被碰掉的电子当然全都是带电粒子，而那些丢了电子的原子——其中大多数已经变成了“一丝不挂”的原子核——也都是带电粒子。因此太阳物质几乎是清一色由带电粒子组成的，是所谓的等离子体（plasma）。光子穿越这种到处都是带电粒子的物质时，就像女人穿越到处都是化妆品专柜的购物城，将会无可避免地受到巨大干扰。

但不管怎么干扰，光子终究是要逃出去的，否则我们就看不到太阳了。我们首先要问的是：在带电粒子的重重包围之下，那些光子究竟是

以何种方式完成胜利大逃亡的庄严任务的？

这个问题的思路并不深奥，甚至在中学物理中就已经有了，因为所谓的光子大逃亡，归根到底是一个能量传输问题，即把太阳核心所产生的能量传输出去。而能量的传输有三种众所周知的方式：传导（conduction）、对流（convection）和辐射（radiation）。从微观上讲，这三种传输方式的差别在于：传导主要是通过物质粒子间的相互碰撞来传输能量，但那些物质粒子本身却并不参与大范围的运动，我们做饭时热量通过锅底传到锅内所用的就是这种传输方式；对流主要是通过物质粒子本身的大范围运动来传输能量，我们做饭时，沸腾的水就是用这种方式在传输能量；而辐射则主要是通过光子本身来传输能量，我们围坐在一个火炉旁所感受到的“温暖”就是用这种方式传输过来的能量。

在这三种能量传输方式中，传导在气态物质中通常起不到主要作用（请读者想一想这是为什么），像太阳那样的巨型“气球”也不例外，因此这一选项可以被排除掉。这样一来，我们就只剩下了两个选项：对流和辐射。一道选择题只有两个选项，这运气看来很不错，因为就算瞎蒙也该有一半的可能性答对。

但有关太阳的这道选择题却是一个例外。

对于这道选择题，早年的天文学家们曾经作出过自己的选择，那就是对流，理由是我们所熟悉的唯一一种大范围气态物质——地球大气——就是用这种方式来传输能量的。

可惜那是一个错误选择。

这一选择在1906年遭到了德国物理学家史瓦西（Karl Schwarzschild, 1873—1916）的反对。史瓦西认为对于像太阳内部那样

与地球大气截然不同的物理环境，很可能存在一种具有太阳特色的能量传输方式，而不能简单地套用地球大气的经验。作为对这一观点的论述，他提出了一种判断太阳内部是否会出现对流的巧妙方法，那就是从理论上分析一小团太阳物质的运动，看它能否演变成对流。

具体地说，史瓦西从理论上考察了一小团由于偶然原因而变得比周围环境稍热的太阳物质。由于热胀冷缩的缘故，那一小团物质的体积会稍稍膨胀，密度则会稍稍降低，由此产生的后果是在周围物质的浮力作用下上升。但在上升的过程中，由于周围太阳物质的压强在变小（因为太阳物质的压强是越往上就越小——请读者想一想这是为什么），它会发生进一步的膨胀，这种膨胀会使它的温度降低。^[2]另一方面，随着这一小团物质的上升，它周围环境的温度也在降低，因为太阳内部物质的平均温度也是越往上（即越远离核心区）越低。既然那一小团物质与它周围环境的温度都在降低，显然就出现了哪个温度降低得更快的问题。如果是周围环境的温度降低得更快，那么该小团物质将会在上升过程中维持比周围环境更热的特点，从而持续上升，那样就会产生对流。反之则它的上升势头会因温度降低得比周围环境更快而终止，并在重力的作用下转为下降，那样就不会产生对流。

这样，史瓦西就提出了一个分析太阳内部能否产生对流的判据，即通过比较太阳内部物质的平均温度变化与一小团特定物质在上升过程中的温度变化，来判断能否产生对流。这一判据被称为史瓦西判据（Schwarzschild criterion）。利用这一判据，史瓦西推翻了前人的选择，代之以自己的选择，那就是辐射。

可惜那也是一个错误选择。

读者可能会纳闷：一道只有两个选项的选择题，怎么会两个选项都

错误呢？原因很简单：因为两个选项都不完全，都只在一定区域内才适用，把无论哪个选项当做完整答案都是错误的。那么，怎样才能得到完整答案呢？只有一个办法，那就是计算。既然史瓦西给出了分析太阳内部能否产生对流的判据，那我们就可以——而且必须——通过计算来作出判断。这种计算所依据的就是我们在前几章中已经多次提到过的标准太阳模型。那个模型是建立在一系列很基本的物理原理——比如能量的守恒、压强的平衡等——的基础之上的，虽然算不上精细，但以粗线条而论却有很大的可靠性。史瓦西本人虽然由于他那个时代的知识所限而没能得到正确结论，但我们这些“站在巨人肩上”的幸运儿却毫无疑问可以做得比他更好一些。

那么利用标准太阳模型所做的计算给出了怎样的答案呢？那答案就是：太阳内部的能量传输方式既不是单纯的对流也不是单纯的辐射，而是在不同区域内有不同的主导方式。具体地说，在从太阳核心区的边缘（即太阳半径的四分之一处）到太阳半径的70%处（即距离太阳中心约49万千米处）的厚度约32万千米的物质层内，史瓦西判据得不到满足，能量的传输以辐射为主。这一层因此而被称为辐射区（radiative zone），它占太阳总体积和总质量的比例分别约为32%和48%。在辐射区的外部边缘，太阳物质的温度降到了“只有”200万度，密度则降到了 0.2克/厘米^3 （相当于水的密度的五分之一）。在那以外直到太阳表面的厚度约20万千米的物质层内，[\[3\]](#)史瓦西判据得到满足，能量的传输方式转为以对流为主，这一层因此而被称为对流区（convective zone），它约占太阳总体积的66%，但由于物质密度很低，在太阳总质量中却只占了2%左右（仍相当于6 600个地球的质量）（彩图8）。

有了这样的大致图像，现在我们可以来谈谈光子大逃亡的具体过程了。由太阳核心核聚变反应所产生的光子大都“膘肥体壮”（满载着能

量），比如第一类质子-质子链所产生的光子的能量在百万电子伏特

（MeV）量级上，属于 γ 射线（gamma ray）的范围。^[4]如果它们在逃亡时有像中微子那样的能耐，接下来的故事将只需两秒钟就能完成，不过那样的话，我们也就不可能在这里读这篇文章了，因为那样的话，地球将会沐浴在致命的 γ 射线之中，生命的产生几乎是不可能的。幸运的是，如前所述，光子由于会参与电磁相互作用，从而在本质上是等离子体的太阳物质中无可避免地会受到巨大阻挠。计算表明，在太阳辐射区中，可怜的光子平均飞行不到1毫米就会遭遇“灭顶之灾”——被带电粒子所吸收。不过带电粒子的胃口也很有限，并无能力把一个高能光子单独“消化”掉。事实上，它们“吞下”光子后一方面会通过碰撞把一部分能量分给其他带电粒子，另一方面会几乎立刻就重新吐出一个或多个光子。那些浴火重生的新光子在能量和运动方向上都有很大的随机性，有些甚至“天堂有路它不走，地狱无门自来投”——重新向着“地狱”（核反应区）的方向飞去。如果我们追踪一个高能光子和它那些被吸收后重新发射出来的“子孙后代”的去向，我们将会发现，它们就像一群醉汉一样随处游荡。

不过，如果我们更仔细地观察这群“醉汉”，我们又会发现一些微妙的特点。比如当它们往太阳表面方向运动时，平均来说会比往相反方向运动时多走一小段距离，那是因为太阳物质的分布是越往表面方向密度越低，因此往表面方向运动的光子在被带电粒子吸收之前就有可能运动较长的距离。由于这个微妙的差别，那些“醉汉”虽然浑浑噩噩，总体上却是在缓慢地向着太阳表面方向运动着。另一个微妙的特点是，被带电粒子发射出来的新光子的“诞生地”越靠近太阳表面，其平均能量就越低。这是太阳物质的温度越靠近表面就越低这一特点所产生的平均意义上的影响。因为这两个微妙特点的共同存在，光子在辐射区中的逃离路线虽然极度曲折，但最终的效果却是慢慢地向外逃逸，而且在逃逸过程

中逐渐“减肥”（能量逐渐降低）。那么，这种逃亡生涯啥时候才是个尽头呢？计算表明，一个高能光子穿越太阳辐射区平均需要17万年的时间，^[5]这并不是因为光子的运动速度变慢了，而纯粹是因为运动路线过于曲折。

在辐射区中，由于史瓦西判据得不到满足，太阳物质不会发生显著的对流，因此这个区域内的太阳物质相对来说是比较宁静的。如果我们能乘坐一艘假想的飞船缓缓穿越这一区域的话，相对来说将不会太颠簸。不过这种宁静在我们抵达辐射区的顶部时就开始消失了。在那里，由于史瓦西判据开始得到满足，太阳物质无法继续维持大体上静态的平衡，无时不在的温度涨落将会使得热气团上升、“冷”气团下降，从而形成对流。如果我们继续乘坐飞船的话，滋味可就不太好受了。

细心的读者在这里可能会提出一个问题：那就是史瓦西判据得到满足只是说明在太阳物质中会产生对流，却并不表明对流一定会成为能量传输的主导方式，那么在对流区里究竟是以哪一种能量传输方式为主导呢？答案是对流（否则就不叫对流区了），因为在对流区里，太阳物质由于温度降低而变得不那么透明了，从而对辐射产生了抑制作用。^[6]在对流区里，光子携带的能量变成了热气团的内能，随着它们的对流运动而传向太阳表面。要说团体的力量还就是大，气团的运动速度虽然远不能与光子相比，但它们却不会像光子那样处处受到阻挠，其结果是，光子要用17万年的时间才能走完厚度为32万千米的辐射区，对流却只需十来天的时间就能走完厚度为20万千米的对流区。

那么，对流区中对流的具体形式，即那些气团的具体运动又是怎样的呢？这却是一个极难回答的问题，因为用流体力学理论所做的简单评估表明，太阳对流区中的对流是所谓的湍流型对流（turbulent

convection），它作为能量传输方式是非常有效的，^[7]但从机理上讲却复杂得令人望而生畏，因为它涉及一种迄今仍未被完全理解的现象——湍流（turbulence）。科学家们研究湍流已有上百年的历史，却始终未能参透它的奥秘。美国物理学家费恩曼（Richard Feynman, 1918—1988）曾把湍流称为是“尚未解决的经典物理学问题中最重要的一个”。而据说著名物理学家海森伯（Werner Heisenberg, 1901—1976）在去世前不久也曾表示，当他见到上帝时将会问上帝两个问题：一个是为什么会有相对论，另一个是为什么会有湍流。他认为上帝有可能可以回答第一个问题。^[8]

但研究太阳结构的科学家们却想要回答第二个问题。

这个愿望迄今仍未实现。不过在长期的研究中，科学家们发展出了一些近似理论，其中很重要的一个近似理论是德国空气动力学家普朗特（Ludwig Prandtl, 1875—1953）在20世纪20年代提出的，它把湍流中的流体元与分子运动论中的分子相类比，尤其是将湍流运动中一个流体元在与其他流体元相混合之前所走过的平均距离与分子运动论中的分子平均自由程（即分子在两次相邻碰撞之间所运动的平均距离）相类比。这种近似理论被称为混合程理论（mixing length theory）。自20世纪30年代开始，德国天体物理学家比尔曼（Ludwig Biermann, 1907—1986）等人将这一理论运用到了研究太阳对流区的能量传输上。20世纪50年代，人们又进一步将这一理论与恒星结构模型结合起来，使之变得更为系统。再往后，随着计算机技术的飞速发展，人们开始对太阳对流区中的运动进行计算机模拟，混合程理论在那里也得到了重要应用。这类理论虽然明显只具有近似意义，但迄今为止的模拟计算显示，它可以给出定性上还算不错的结果。这种类型的结果是当前人们对太阳对流区研究的最佳成果之一。

与深藏在太阳内部的核反应区及辐射区不同，对流区由于一直延伸到太阳表面，因此对这一区域的研究结果是可以与直接观测相对比的。这就好比当我们研究一锅沸水的时候，哪怕这水是不透明的，我们也可以通过观察因水流翻腾而形成的表面波纹来推断水的沸腾程度及其他一些相关特征。太阳的表面也是如此。早在1801年，英国天文学家赫歇耳（William Herschel, 1738—1822）就在太阳表面——确切地说是在太阳大气层底部的光球层中——发现了无数形如米粒的斑点，这种斑点被称为米粒（granule）——科学家们在取名字方面的想象力看来是很贫乏的。

这些“米粒”当然只是远远看起来才像是米粒，如果跑近了看的话，每个的线度都有1—500千米左右，是面积比新疆还大的巨型结构，只不过是放在太阳的庞大表面上，才显得微乎其微而已。简单的估算表明，太阳表面足可容纳几百万个“米粒”。彩图9是一幅典型的“米粒”图片，细致的观测表明，在那些“米粒”中，炽热的气体以每秒几百米的速度从明亮处升起，并向外迸射出大量光子——它们此刻的能量已降到了可见光区附近。由于那儿已是太阳的表面，再没有什么东西可以有效地阻碍重获自由的光子，它们随即以天下第一的逃命速度飞离太阳，为历时十几万年的长征画上句号。由于光子的逃离带走了能量，“米粒”中的气体温度迅速下降，并在颜色较暗的边缘部位收缩下沉。一个“米粒”的典型寿命只有几分钟。

太阳表面的“米粒”是发生在对流区中的对流运动的体现，这一点是毫无疑问的。但与这种对流有关的一个插曲值得在这里介绍一下。曾经有一段时间，天体物理学家们普遍认为“米粒”以及太阳表层的对流运动是人们在流体力学中所发现的一种称为贝纳胞（Bénard cell，彩图10）的结构在太阳上的翻版。贝纳胞是1900年法国物理学家贝纳（Henri

Bénard, 1874—1939) 在自下而上加热一个液体薄层时所发现的对流形式，它与太阳表面的“米粒”不仅具有外观上的相似性，而且在生成条件上也一度被认为是相同的，因为两者都被认为是在重力和浮力的共同作用下由对流运动产生的。

粗看起来，这是一个展示不同尺度物理现象之间相似性的漂亮结果。但这种相似性却在1958年遭到了严重质疑，因为人们发现在贝纳胞的形成过程中，一个曾经被忽略的因素——液体的表面张力——起了不可忽视的作用。为了证实这一点，科学家们甚至在“阿波罗14号”（Apollo 14）登月飞船的无重力环境下进行了贝纳胞实验，结果发现在重力和浮力都不存在的环境下依然可以出现贝纳胞，它的幕后推手是表面张力。这一结果不仅打破了对贝纳胞的传统理解，而且直接削弱了它与太阳“米粒”之间的可比性，因为在太阳表面极端稀薄的气体环境里并不存在表面张力，因此“米粒”的形成机制与贝纳胞起码是不可能完全相同的。这段插曲是一个很好的例子，说明在太阳物理学的研究中存在很多微妙的地方，只有很仔细地考虑到太阳环境的特殊性，才能避免似是而非的理解。

太阳表面的“米粒”除了那无数个线度约为1 500千米的“小”米粒外，20世纪50年代，人们在观测太阳表面物质的速度分布时，还发现了一种线度为2万~3万千米的大尺度结构，它的面积比整个地球的表面积还大，它名字则跟“米粒”一样贫乏，叫做“超米粒”（supergranule）。这种“超米粒”的寿命比普通“米粒”长得多，可以有1~2天。在太阳圆面上，几乎每一时刻都分布着两三千个“超米粒”。与“米粒”不同的是，“超米粒”的速度分布基本上是水平的，无法找到像热气团上升、“冷”气团下沉那样简单的物理结构，从而无法把它们与对流运动简单地联系起来。“超米粒”的结构和成因也因此而成为了太阳物理中的一

个不解之谜。

更热闹的是，除了“米粒”和“超米粒”外，1981年，人们还发现了一种线度约几千千米，介于“米粒”和“超米粒”之间的新型“米粒”，它的名字读者们一定猜到了，就叫做“中米粒”（mesogranule）。“中米粒”的发现受到了一部分天体物理学家的欢迎，因为早在1961年就有人提出过，太阳对流区中的对流有可能具有不同的层次，“米粒”和“超米粒”分别对应于其中较小和较大的层次，但介于两者之间的层次却一直未能得到观测上的支持。“中米粒”的发现可以说是填补了这一空白。但不幸的是，这一观测与理论之间看上去很美的契合后来也遭到了质疑。有些科学家在仔细研究了观测数据后，认为“中米粒”有可能只是对太阳表面物质的速度分布进行分析时采用了不恰当的统计方法所导致的“幽灵结构”，它其实根本就不存在。这方面的争议迄今仍未得到解决。

因此，不仅“超米粒”是一个谜，“中米粒”也是一个谜，只不过前者之谜在于结构和成因，而后者则干脆连存在性本身都是一个谜。那么普通“米粒”又如何呢？它是唯一一种与对流运动存在明显对应的“米粒”，从这个意义上讲，它是三种“米粒”当中被理解得最充分的一种。但即便对于它，也仍有很多微妙的地方有待理解，否则人们也就不会轻易犯下将它视为贝纳胞那样的错误了。事实上，如果把太阳上各种主要的复杂因素——比如太阳的自转或磁场等——的影响全都考虑在内的话，即便对于普通“米粒”的结构和成因，也还有很多需要探索的地方。

现在让我们把迄今介绍过的太阳内部结构列成一个简单的表格：

区域名称	范 围	主要现象	检验方法
核心区	0~17万千米	核聚变反应	探测各个能区的太阳中微子

辐射区	17万~49万千米	以辐射为主的能量传输	?
对流区	49万~69.5万千米	以对流为主的能量传输	观测太阳表面的各种“米粒”

这个表格最引人注目的地方想必读者们都看出来，那就是在辐射区的检验方法一栏中有一个问号。其实不仅在辐射区的检验方法一栏中有一个问号，在对流区的检验方法一栏中也起码有半个问号，因为观测太阳表面的各种“米粒”虽然能告诉我们一些有关对流区的信息，但那些信息大都局限在最靠近太阳表面的那部分对流区里，相对于厚度达20万千米的整个对流区来说充其量只能算是“皮毛”。

那么，有没有什么手段，能像太阳中微子带给我们核心区的信息那样，带给我们有关对流区深处及辐射区的信息呢？

[1]当然，太阳核心区的边界并不是一个鲜明的几何边界，而是一个有一定厚度的渐变区域，而且即便在这一渐变区域之外，核聚变反应也只是少到可以忽略而并非绝对不可能。这一点不仅适用于核心区的边界，而且也适用于后面将要提到的任何其他区域之间的边界。

[2]在史瓦西的分析中，这一小团物质的上升过程被假定为是绝热的，即与周围太阳物质之间几乎不交换热量。这是一种近似。在这种近似下，膨胀会使这一小团物质的内能减少（因为一部分内能会消耗在膨胀导致的对外做功上），从而使得温度降低。另外顺便提一下，这位史瓦西就是那位给出了广义相对论的第一个非平凡严格解——史瓦西解——的史瓦西。

[3]需要提醒读者注意的是，这里所说的太阳表面并不是我们肉眼所见的太阳表面，后者——如我们将在后文中加以介绍的——其实是隶属于太阳大气层的。

[4]人们通常把能量在0.1 MeV以上的光子称为 γ 射线。作为对比，可见光光子的能量只有2~3 eV。

[5]熟悉布朗运动 (Brownian motion) 或随机游走 (random walk) 问题的读者可以利用本文提供的信息自己估算一下这一时间，看能否得出数量级相近的结果。

[6]温度降低之所以能使太阳物质变得不那么透明，一个重要的原因是随着温度降低，重原子核开始有机会俘获一部分电子，这些俘获了一部分电子的重原子核吸收光子的能力比裸核大得多（因为它们有密集的电子能级，可以通过电子跃迁吸收光子）。太阳物质在对流区里变得不那么透明不是偶然的，事实上，正是因为太阳物质变得不那么透明了，才使得温度分布发生了新的变化，进而使史瓦西判据得到了满足。

[7]湍流性对流不仅是一种极为有效的能量传输方式，而且对太阳来说还有许多其他影响，比如能使太阳的物质均匀化，以及能把太阳深处的重元素带到太阳表面等。人们能够在太阳光谱中发现重元素的谱线，在很大程度上就得益于湍流性对流的存在。

[8]这个故事流传甚广，但可信度不高，八卦价值大于历史价值。



绘画：张京

12 太阳的脉搏

为了回答第11章末尾的问题，即“有没有什么手段，能像太阳中微子带给我们核心区的信息那样，带给我们有关对流区深处及辐射区的信息呢”？让我们把时钟拨到1960年。那一年，在意大利科摩湖畔的一座美丽小镇召开了一次天文学会议。在会上，来自美国威尔逊山天文台（Mount Wilson Observatory）的天文学家莱顿（Robert Leighton, 1919—1997）作了一个学术报告。

这位莱顿，学物理的读者可能有点印象，他曾与同事一起，用两年时间，将著名美国物理学家费恩曼（Richard Feynman, 1918—1988）的讲课录音整理成风行全球的《费恩曼物理学讲义》（*The Feynman Lectures on Physics*）。在1960年的那次报告中，他介绍了自己对太阳大气层中气流运动的观测研究。

这不是一项轻而易举的研究。我们知道，在地球大气层中如果要观测气流运动，最简单的办法就是放置风速风向仪，但这招显然无法用于太阳。那么，太阳大气层中的气流运动该如何观测呢？科学家们想到了一种物理效应：多普勒效应（Doppler effect）。该效应的一个让人耳朵听出老茧来的例子，是火车交汇时对方火车的汽笛声由迎面而来时的尖锐，变成交错而过后的低缓。用物理学家们的术语来说，多普勒效应显示的是波源与观测者的相对运动对波长和频率的影响，它既适用于声波，也适用于光波，只不过后者的波速实在太快，使我们无法用感官直接体验。但科学家们可以用仪器来延伸自己的感官，从而可以观测光波的多普勒效应，并以此推算出光源相对于我们的运动速度。

观测光波多普勒效应的最典型做法，是对光波的光谱进行精密观测。我们在第6章中介绍过，每种元素都有自己独特的光谱，就像每个人都有独特的指纹一样。而所谓独特的光谱，说白了就是独特的谱线波长和频率。因为有这种独特性，当谱线的波长和频率因多普勒效应而发生偏移时，我们就能明确无误地察觉出偏移，并依据偏移大小推算出光源的运动速度，这正是莱顿所用的基本方法。^[1]

通过这种方法，莱顿发现了什么呢？他发现太阳这个庞然大物在颤抖，而且是有规律地颤抖！确切地说，他发现太阳大气层中的气流运动不是完全无序的，而是存在一种周期约为五分钟的振荡，这种振荡被称为“五分钟振荡”（five-minute oscillation）。

这一发现在太阳研究中具有里程碑式的意义，虽然其真正价值直到十几年后才被发掘出来，但它当场就给与会者中的一位带来了巨大震动。此人名叫埃文斯（John Evans, 1909—1999），是萨克拉门托峰天文台（Sacramento Peak Observatory）的台长。莱顿的报告之所以给他带来巨大震动，是因为他研究的恰好也是太阳大气层中的气流运动。为什么莱顿发现了新东西而他却没有呢？因为他的研究重点是气流运动的空间分布，而非时间变化。平心而论，他对研究重点的这一选择并非毫无道理，因为如我们在第11章中所说，太阳上的气流运动是所谓的湍流型对流，这种运动的时间变化被认为是无序的，空间分布则因为有“米粒”之类的结构，以及物质密度的逐层递变，而不无探索余地。因此，埃文斯的研究重点可以说是很自然的选择。遗憾的是，科学发现有时偏偏是出人意料的，“五分钟振荡”对埃文斯来说就是如此。他对研究重点的自然选择恰恰使他与该领域最重大的发现失之交臂。

但遗憾归遗憾，成熟的研究者是不会因遗憾而消沉的。“五分钟振

荡”虽已被发现，进一步的观测仍是必不可少的。更重要的是，这一现象的产生机制还是一个空白，而且在这点上，莱顿本人表示了一定程度的悲观，他认为对“五分钟振荡”做出精确计算似乎是毫无希望的。这一切对于其他研究者来说无疑都是机会。

回到萨克拉门托峰天文台后，埃文斯立刻对“五分钟振荡”展开了研究。第二年，莱顿与埃文斯又在学术会议中相遇了。这一回，埃文斯也作了有关“五分钟振荡”的报告。在报告中，他不仅证实了莱顿的观测，还利用自己在以前那些与“五分钟振荡”失之交臂的研究中发展起来的“独门绝活”，对莱顿的观测作出了重要补充。他的“独门绝活”是什么呢？是同时研究几条不同谱线的多普勒效应。这种“独门绝活”有什么用呢？用处就在于能研究太阳大气层中不同高度处的气流运动。这是因为不同高度处的元素分布存在一定差异，而谱线是由元素产生的，因此不同谱线所对应的是不同高度处的气流运动。通过对这种不同高度处的气流运动的观测，埃文斯对“五分钟振荡”在太阳大气层中的空间传播进行了粗略研究，并提出了一种“五分钟振荡”的产生机制。

埃文斯认为，“五分钟振荡”是一种由太阳表面的“米粒”所激发的大气层现象。具体地说，他认为“米粒”的上升犹如推动气体的巨型活塞（piston），而“五分钟振荡”则是“活塞”运动在太阳大气层中激发出的声波。这一机制听起来不无道理。因为“五分钟振荡”是在观测太阳大气层中的气流运动时发现的，而“米粒”，如我们在第11章中所说，则是在太阳大气层底部的光球层中被发现的，而且是光球层中分布最广的结构。这两者同属太阳大气层，将它们联系起来，并在太阳大气层的范围之内解决问题无疑是很自然的思路。这种很自然的思路在将近十年的时间里吸引了多数研究者的目光。

不过，埃文斯的机制要想行得通，还必须解决一个关键性的问题，

那就是被“活塞”激发出的声波为什么偏偏要青睐“五分钟”这一振荡周期呢？这个问题成为了很多研究者的努力方向。

在介绍那些努力之前，我们要先对太阳大气层的结构作一个简单介绍。如上所述，太阳大气层的最底部，是“米粒”们赖以存身的光球层（photosphere）^[2]。我们肉眼所见的阳光大都来自该层，第7章中介绍过的太阳表面温度（约5 800 K）也是指该层的有效温度。该层的厚度在几十到几百千米左右，物质密度约为海平面附近地球大气密度的万分之一。在光球层之上，是所谓的温度最低层（coolest layer），该层的有效温度约为4 100 K，厚度约为500千米。再往上，则是厚度约2 000千米的色球层^[3]。该层自内向外密度递减几百万倍，温度却不降反增，上层温度可达20 000 K左右（原因将在后文中介绍）。色球层再往上，还有所谓的过渡区（transition region）和日冕（corona）等，因与本章无关，暂且按下不表。

现在回到埃文斯机制所面临的那个关键问题上来。很明显，太阳表面环境的恶劣性，使得无论什么机制产生的声波都必然是紊乱的，而不可能只有一个固定周期。那么，究竟是什么原因使得“五分钟”这一周期脱颖而出呢？科学家们想到了乐器中的谐振腔（resonant cavity）。在乐器中，谐振腔的作用是使特定周期的声波得到加强，其余则被抑制。科学家们想到，如果太阳大气层也有类似功能，岂不就有可能解释“五分钟振荡”了吗？

这是一个很漂亮的想法。不过科学离不开细节，光有想法是不够的，还必须有计算。为了进行计算，我们必须知道太阳大气层是如何构成谐振腔的？我们知道，谐振腔之所以能加强特定周期的声波，抑制其余，是因为声波经腔壁反射后与腔内声波发生干涉之故。因此，要想知

道太阳大气层如何构成谐振腔，关键在于找到腔壁。对此，科学家们提出了各种假设。

比如有一种假设认为，腔壁位于温度最低层的上下两侧。为什么呢？因为气体中的声波波速会随温度升高而增大，^[4]这意味着在温度最低层的上下两侧，声波波速都会增大（因为温度升高）。另一方面，当声波从波速小的区域进入波速大的区域时，会因折射而往水平方向偏折。^[5]将这两点联系起来，就有可能出现这样一种情形，即在温度最低层附近的区域里，声波无论向上还是向下传播，都会往水平方向偏折，直至被反射回来为止。这种情形一旦成真，那两个使声波反射回来的界面无疑就构成了谐振腔的腔壁。

类似的假设还有若干种，我们就不一一列举了。可以告诉大家的是，所有这些在太阳大气层里做文章的假设在经过具体计算后，全都遭遇了滑铁卢。埃文斯是一位努力的研究者，但看来不是一员福将，他对研究重点的“自然选择”使他与“五分钟振荡”的发现失之交臂，他对产生机制的“自然选择”则将一系列努力引向了困难。这其中一个典型的困难是因为太阳大气层内的谐振腔全都是薄层，声波要想被那样的薄层所禁锢，其传播方向必须很接近水平（否则的话，虽然能被偏折，却无法偏折到被反射的程度）。而观测却发现，“五分钟振荡”的传播方向有很大的垂直分量。另一方面，即便声波的传播方向真的很接近水平，那些假设也还是玩不转，因为计算出的周期不对头，无法集中在五分钟附近。此外，有些假设给出的“五分钟振荡”的寿命也不对，只有几分钟，而实际观测到的“五分钟振荡”往往可以持续数小时、数天、乃至数月。至于埃文斯把“米粒”视为“五分钟振荡”的起因，其正确与否倒显得不那么重要，因为在太阳那样的恶劣环境下，有许多因素可以产生声波，起因不是关键。

所有假设都碰了壁，莫非应了莱顿的“乌鸦嘴”，对“五分钟振荡”做出精确计算是毫无希望的？

没有人知道答案。但任何谜团对科学家来说都是召唤，是不断攀登的动力。山峰越高，顶上的风景或许就越别致。1970年，一位新的攀登者——加州大学洛杉矶分校（University of California at Los Angeles）的一位助理教授——提出了一种“五分钟振荡”的新机制。这位助理教授名叫乌尔里克（Roger Ulrich），本科学的是化学，后来转向天文，1968年才刚刚获得博士学位。这位“新科博士”的过人之处是一举摆脱了在太阳大气层中做文章的“陋习”。对乌尔里克来说，“五分钟振荡”在太阳大气层中被发现，并不意味着它一定就是单纯的大气层现象。理由很简单，太阳物质在大气层以下是不透明的，因此普通光学手段注定只能发现大气层中的运动，这只是观测手段的局限性，并不说明现象本身的范围。这就好比用肉眼只能看到水面以上的冰山，并不说明冰山只存在于水面以上。

事实上，在乌尔里克之前，美国基特峰国家天文台（Kitt Peak National Observatory）的天文学家弗雷泽（Edward Frazier）就已经发现了“五分钟振荡”存在于太阳深处的迹象。因为他发现太阳的亮度会因“五分钟振荡”而发生细微变化。由于太阳的亮度取决于能量，而后者来自太阳内部，单纯的大气层振荡是无法对其产生重大影响的，因此亮度变化说明“五分钟振荡”与太阳内部物质的运动也有关系。弗雷泽的这一观测发现给了乌尔里克很大的启示，使他将注意力投向了太阳内部。

乌尔里克将注意力投向太阳内部，除受弗雷泽的影响外，与他自己的学术背景也不无关系，因为他博士论文所研究的就是太阳内部的对流区。方向有了，接下来还得看细节。乌尔里克知道，那些被观测否决掉的大气层假设并非一无是处，它们的一个关键部分必须得到保留，那就

是谐振腔，因为那是在像太阳那样高度无序的世界里产生特定周期的唯一手段。既然保留了谐振腔，那么新机制的关键就依然是：腔壁在哪里？对于这个问题，乌尔里克给出了全新的答案。

在乌尔里克的答案中，上方腔壁位于对流区顶部很接近光球层的区域。该区域之所以成为腔壁，是因为声波在太阳内部的传播存在一个所谓的截止频率（cut-off frequency），只有频率高于截止频率的声波才能传播。不仅如此，这一截止频率还与温度有关，温度越低（即越接近太阳表面），截止频率就越高，能够传播的声波也就越少。^[6]计算表明，在对流区顶部接近光球层的区域中，截止频率将会高到使几乎所有声波都无法传播的程度。自内向外的声波碰到这种区域时，将会像光线碰到镜面一样遭到反射^[7]，这毫无疑问就构成了腔壁。

那么，下方腔壁又在哪里呢？在太阳肚子里。这一腔壁的原理（彩图11）与上文介绍的大气层假设相同，即源于声波的折射。具体地说，越往太阳肚子里走，声波的波速就越高（因为温度升高），经折射后就越往水平方向偏折，直至被反射回来为止。与大气层假设不同的是，乌尔里克的机制将整个太阳内部都作为声波的活动场所，有足够大的空间使声波“浪子回头”，因而不必像大气层假设那样要求其传播方向接近水平。另外，从这一机制中不难看到，下方腔壁的位置不是固定的，而与声波的传播方向有关。传播方向越接近水平（比如图中位于对流区中的红色曲线），下方腔壁的位置就越浅，声波在上方腔壁的反射次数则越多（即水平方向的波长越短）；反之，传播方向越接近垂直（比如图中透入辐射区的黄色和绿色曲线），下方腔壁的位置就越深，声波在上方腔壁的反射次数则越少（即水平方向的波长越长）；当传播方向垂直到一定程度时（比如图中深入核心区的蓝色曲线），甚至会出现下方边界消失，完全靠上方边界反射的情形（这依然构成谐振腔）。

定量计算表明，乌尔里克的机制可以克服大气层假设所面临的各种困难。不仅如此，这一机制还预言了一些尚未被观测到的东西。其中最突出的一个，是它预言“五分钟振荡”不是单一频率的振荡，而是由一系列彼此接近的频率共同组成的，每个频率对应于一个不同的水平波长（horizontal wavelength）。如果将频率与水平波长间的关系绘成图线的话，很像是一系列抛物线。这个预言对乌尔里克机制具有判决性的意义。如果它被证实，那么乌尔里克机制就会像科学史上许多其他成功理论一样，经由从观测到理论，从理论到预言，再从预言到证实那样的典型模式而被确立。反之，如果它被否决，那就没啥可说了，该机制只能乖乖入住博物馆。

那么，观测给出了怎样的判决呢？别着急，得等五年时间，因为当时人们对“五分钟振荡”的观测精度还不够高。在等待期间有必要提到的是，几乎与乌尔里克同时，两位哈佛大学（Harvard University）的研究者也提出了类似的机制，并进行了数值计算。不过他们没有预言频率与水平波长间的关系。1972年，另一位研究者对乌尔里克的机制进行了改进，去除了一些粗糙近似。1975年，日本东京大学（Tokyo University）的两位研究者对声波在上方腔壁的反射进行了更细致的研究，给出了频率与水平波长间更精确的关系（与乌尔里克的结果相差不大）。

与这些理论进展同步，观测数据也在积累之中。乌尔里克本人无疑是最急切盼望观测结果的人，为了让观测“提速”，他将自己的研究生罗兹（Edward Rhodes）派往了萨克拉门托峰天文台（即埃文斯工作过的地方）。1975年，当罗兹积累了足够多的数据，准备撰写博士论文时，一个不幸的消息传了过来：他被人抢了——不是抢劫，是抢先。德国夫琅禾费研究所（Fraunhofer Institute）的天文学家德伯纳（Franz-Ludwig Deubner）已经发表了类似的观测结果。一年后，罗兹也发表了自己的

结果。那些观测结果没有让乌尔里克失望，它们的精度虽只能达到百分之几的量级，却明显地显示出了乌尔里克预言的抛物线（图12.1）。

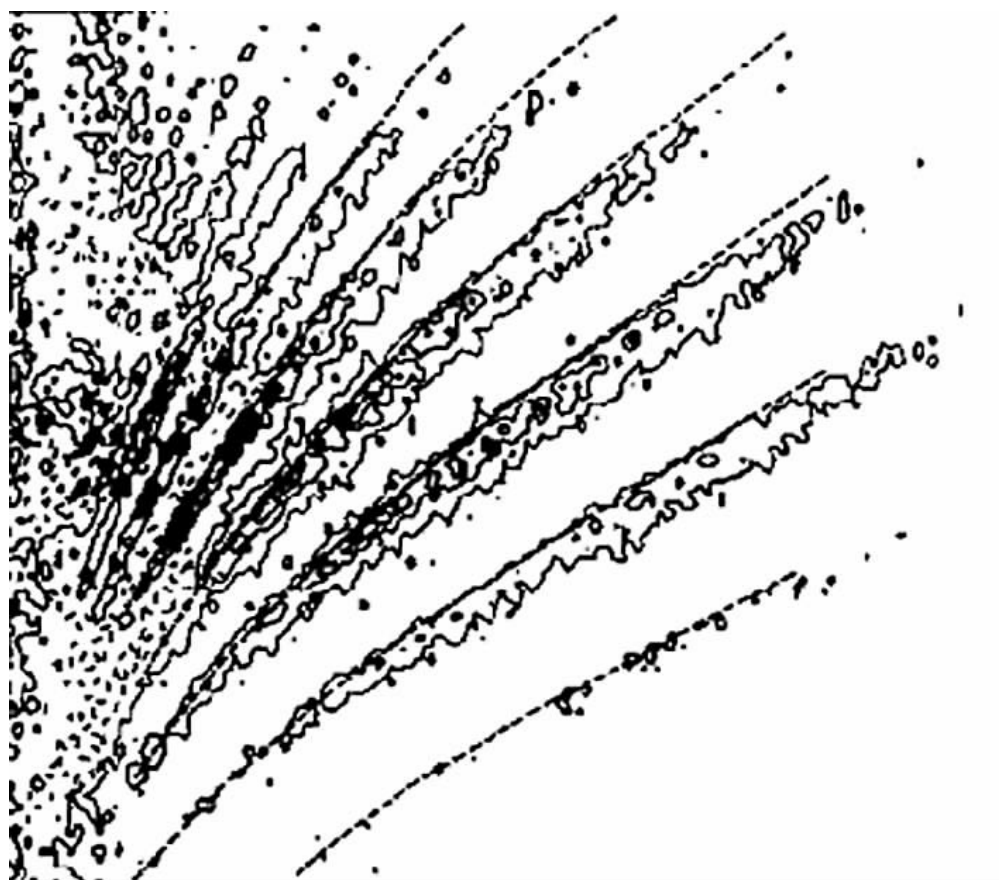


图12.1 “五分钟振荡”的早期观测结果

这样，乌尔里克的机制就得到了初步证实。这不仅对乌尔里克是一个喜讯，对于太阳研究更是具有划时代的重要性，因为它开启了一个全新的研究领域：日震学（helioseismology）这是一个通过观测太阳上的各种振荡现象来研究太阳内部结构的新领域。在那些振荡现象中，以类型而论，最重要的是上面所介绍的声波，这是一种压强波，也称为p模（p来自于“压强”的英文pressure）；以周期而论，最重要的是上面所介绍的“五分钟振荡”。但除此之外，也还有其他一些类型和其他的周期，不过因为观测比较困难，重要性相对较低，就不在这里讨论了。[\[8\]](#)日震学

之所以重要，是因为乌尔里克机制告诉我们太阳上的振荡可以深入太阳内部的各个区域，因此日震学可以带给我们有关太阳内部各个区域的重要信息。像温度、密度、元素丰度、自转速度等与振荡传播有关的性质都可以通过日震学手段来加以研究。打个比方来说，太阳上的振荡仿佛是太阳的脉搏，而日震学手段则相当于是给太阳把脉问诊。乌尔里克攀到了山顶，那里的风景果然很别致，可以一直看到太阳的肚子里。不过要指出的是，日震学这个与地震学相类似的名称其实并不贴切，太阳上虽然有日震，但日震只是产生振荡的诸多原因之一，而且日震本身并非日震研究的对象。

日震学这一研究领域虽然诞生了，但为了让它真正发挥作用，除不断改进理论计算外，还必须在观测上精益求精，其中最重要的一条是得把振荡周期尽可能测准了。在大学甚至中学物理实验课上做过周期测量的读者想必都知道，要想把一个周期运动的周期尽可能测准，要诀是观测尽可能多的周期数目。周期数目越多，所测周期的相对误差就越小。那么，为了让日震学真正发挥作用，需要观测多少个周期呢？答案是一万个以上。对于“五分钟振荡”来说，一万个周期约为35天，因此需要连续观测35天以上。

粗看起来，这似乎不是什么大不了的要求。科学研究花费几个月、几年甚至更长时间都是家常便饭，我们在第9章中介绍过的探测太阳中微子的实验就动辄持续几年以上。但对“五分钟振荡”的观测却有一个问题，那就是太阳只有在晴朗的白天才能被观测到，一到晚上就没有了，而光子又不像中微子那样能够穿透整个地球。因此连续35天以上的观测绝不是轻而易举的事情，非但不轻而易举，经这么一说，简直要变成“不可能任务”了。幸运的是，情况也没那么糟，因为地球上有些地方的太阳是可以连续半年悬挂在天上的，那就是两极附近的地区，那里有

所谓的极昼（polar day）。

看来，要想对“五分钟振荡”进行精密观测，必须到极地去。极地的环境那是出了名的恶劣，温度低就不用说了，最麻烦的是，到了极地也不一定就能进行连续观测，因为极昼不等于晴天，冰雾（ice fog）、风暴（storm）、卷云（cirrus cloud）等都会造成观测的中断。而且极地的生活条件那是相当的艰苦，通信联络那是相当的不便。到那里去做研究，套用一句网络流行语来说，叫做“哥做的不是研究，而是寂寞”。即便如此，依然有一些天文学家心甘情愿地去过寂寞日子。1980年1月，两位法国天文学家格雷克（Gerard Grec）和福塞特（Eric Fossat）赴南极进行了观测，可惜运气一般，只获得了5天的连续观测时间。不久之后，美国基特峰国家天文台的天文学家哈维（Jaack Harvey）和杜瓦尔（Tom Duvall）也去了南极，在一个小得像清洁间的房间里一住就是两个月。不仅如此，他们两人后来还多次重返南极，其中有一次获得了长达65天的连续观测时间。这些研究为日震学的崛起立下了汗马功劳，也再次并且非常漂亮地证实了乌尔里克的那些抛物线（图12.2）。

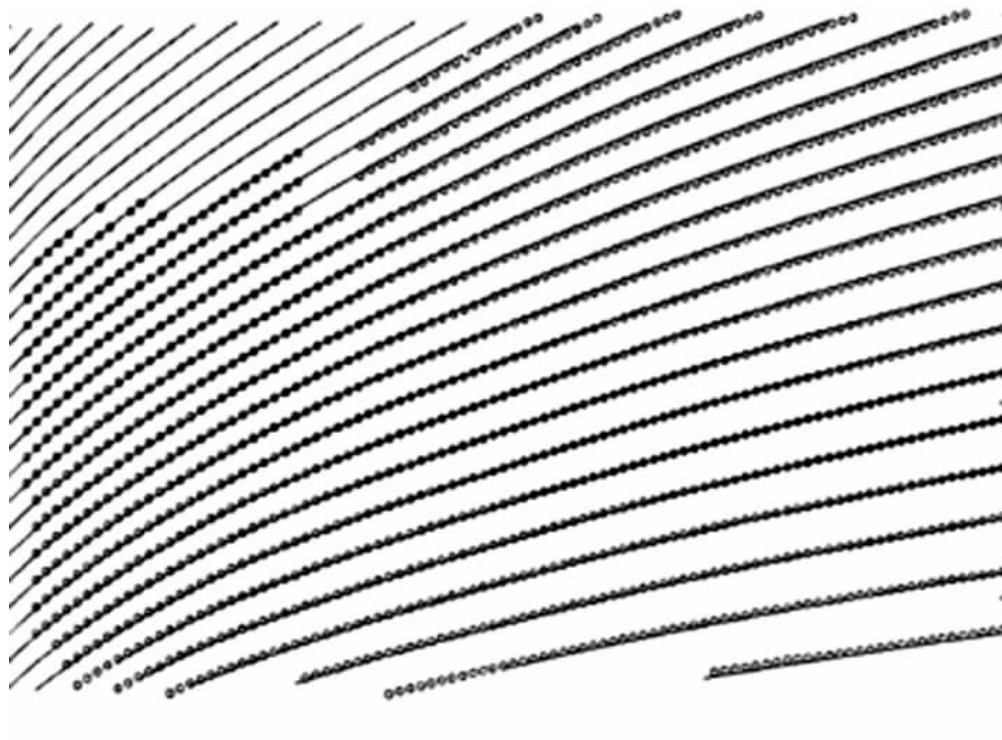


图12.2 “五分钟振荡”的南极观测结果

利用极昼并不是获得长时间连续观测的唯一手段。熟悉历史的读者想必知道，当年英国殖民地遍布全球时曾号称“日不落帝国”，因为太阳在任何时候总能够照到它的某一块殖民地上。与这种帝国的原理相类似，我们也可以用分布于世界各地的若干观测站来构筑一个“日不落”观测系统。1979年，英国伯明翰大学（University of Birmingham）的科学家率先通过两个观测站实践了这种方法。自20世纪90年代中期开始，美国牵头组建了一个更大的“日不落”观测系统，由位于加那利群岛、西澳大利亚、美国加州、美国夏威夷州、印度及智利的六个观测站组成，称为太阳全球振荡监测网（Global Oscillation Network Group, GONG）。

与太阳全球振荡监测网的建设几乎同时，一种更优越的观测手段也开始付诸实施，那就是人造卫星。这种手段的优势是不言而喻的，建在地面上的观测站，无论是两极附近的观测点，还是“日不落”的观测网，都免不了要“看天吃饭”。只有翱翔于九天之上的卫星，才能真正自由自

在地对太阳进行连续观测。1995年，美国国家航空航天局（NASA）发射了一颗“太阳和日球层探测器”（Solar and Heliospheric Observatory, SOHO）。该探测器定位在太阳与地球之间的所谓第一拉格朗日点^[9]（first Lagrangian point, L_1 ）上，与地球公转同步地绕着太阳运转，可以常年不断地观测太阳。利用这样的现代手段，对“五分钟振荡”的观测精度已经达到了百万分之一以上的量级。

那么，日震学研究取得了什么成果呢？最重要的成果就是对标准太阳模型进行修正及检验。我们在前文中提到过，太阳模型看似粗糙，实际上却相当精密。之所以如此，除了它对物理原理的运用相当缜密外，另一个很重要的原因，就是标准太阳模型中的参数大都经过了日震学手段的检验及修正^[10]。

日震学研究所取得的另一项重要成果，是对太阳内部自转速度的分布作出了精密推断。与像地球这样的固态星球不同，太阳内部不同深度、不同纬度处的自转是不同步的。日震学研究不仅发现并测定了这种不同步性，而且还发现这种不同步性在对流区底部以下的一个被称为差旋层（tachocline）的薄层中，非常突然地转变成了刚性自转。这一点是由前面提到的罗兹和德伯纳共同发现的，这两人因验证乌尔里克的预言而“撞车”后，很快化“敌”为友成为了合作者，联手展开新的研究。他们所发现的这个差旋层虽然很薄，而且埋藏得很深，却很可能有着极大的重要性。一般认为，发生在差旋层中的这种自转突变现象与太阳磁场的产生有可能存在密切关联，但很多细节仍有待进一步研究。有读者可能会问：太阳的自转方式为什么会在差旋层那样一个薄层内发生突变呢？很遗憾，这是有关太阳的诸多谜团中的一个，迄今尚无答案。

除上述成果外，日震学研究还有一个引人注目的方面，那就是能够

预报所谓的太阳活动（solar activity）——即诸如黑子（sunspot）和耀斑（flare）那样的现象。那些现象有时会对地球造成影响，比如耀斑有时会干扰无线电通信。日震学为何能预言太阳活动呢？因为它可以探测到太阳背面的活动。它为何能探测到太阳背面的活动呢？因为那些活动大都跟太阳磁场有关，而太阳磁场会使太阳物质产生压缩、沉降等作用，那些作用又进而会影响到声波的传播，使得从太阳正面传往背面，并经背面反射回正面（整个过程约需6~7小时）的声波比正常情况提前十来秒钟。通过对这一点的观测，天文学家们能相当准确地描绘出太阳背面的活动。由于太阳自转一圈需要二十几天，因此描绘出太阳背面的活动，意味着天文学家们有一定的可能性（请读者想一想，为什么只是可能性？）提前若干天预言那些转到正面后会对地球产生影响的太阳活动。这或许是日震学研究中最具实用意义的一面。

最后，让我们将第11章末尾的表格补全，作为对当时所留问题的回答：

区域名称	范 围	主要现象	检验方法
核心区	0~17万千米	核聚变反应	探测各个能区的太阳中微子及利用日震学手段
辐射区	17万~49万千米	以辐射为主的能量传输	利用日震学手段
对流区	49万~69.5万千米	以对流为主的能量传输	观测太阳表面的各种“米粒”及利用日震学手段

[1]感兴趣的读者请思考这样一个问题：当我们认为一条谱线的“波长和频率因多普勒效应而发生偏移”时，如何确定它到底是哪条谱线的波长和频率发生了偏移？又如何排除另外的可能性，比如波长和频率未发生偏移的新谱线呢？

[2]有少数文献将太阳大气层定义为光球层以上的部分，不含光球层。

[3]有些文献将温度最低层视为色球层的一部分。

[4]太阳物质中的声波波速大致与温度的平方根成正比。

[5]确切地说是往两个区域的分界面方向偏折，但由于实际计算中太阳物质的分布被视为标准的层状分布（比如同心球面分布），分界面方向与太阳表面方向基本一致，因此称为水平方向。

[6]太阳物质中的声波截止频率大致与温度的平方根成反比。之所以存在截止频率，是因为当声波的频率太低（即振动太缓慢）时，太阳物质将会有足够的时间自我调节，以抵消声波造成的扰动，从而使声波迅速衰竭。

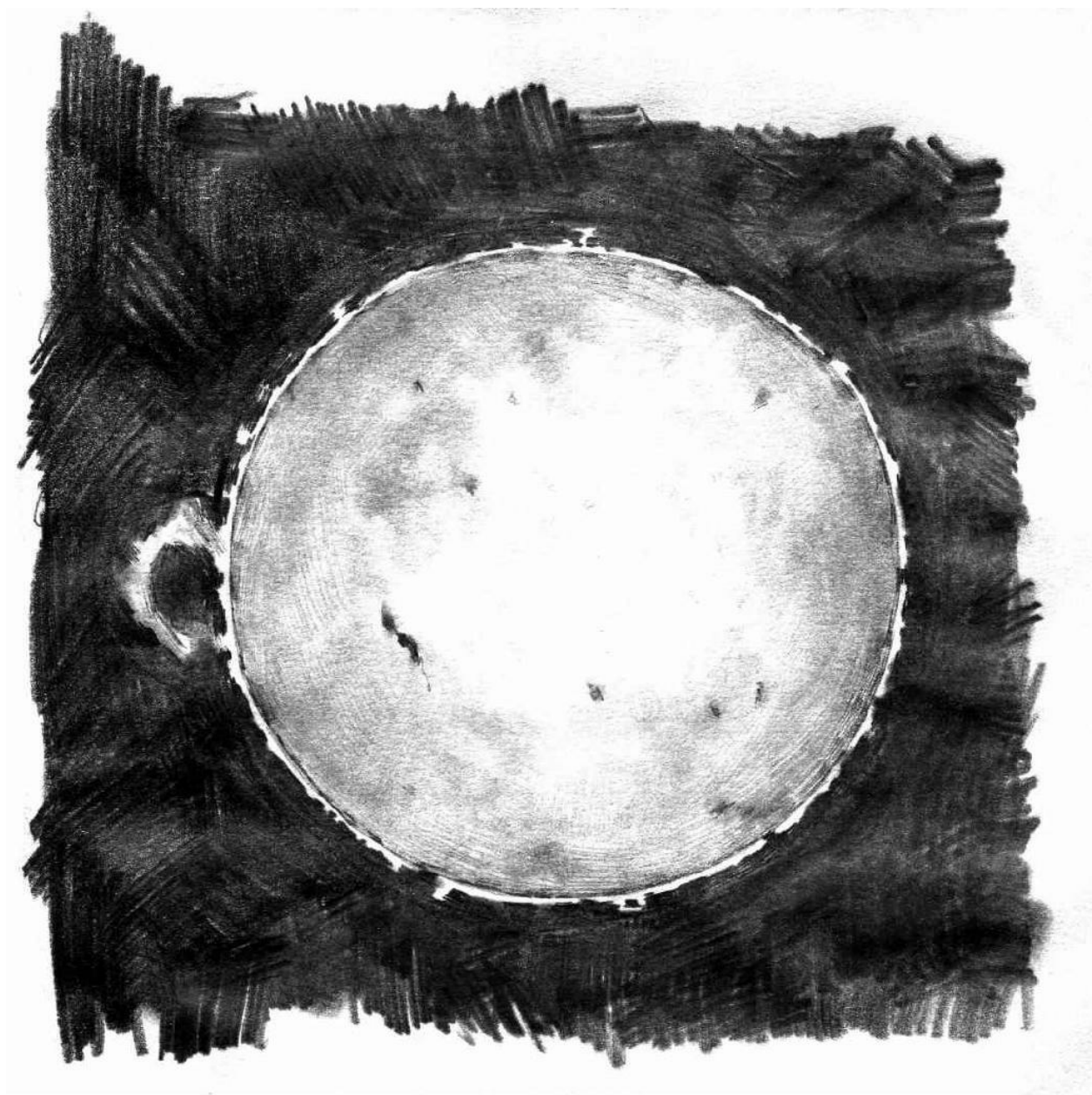
[7]当然，这是理想情况，声波在对流区顶部的实际反射不是百分之百的，而是有一部分会“漏网”，人们在大气层中观测到的振荡就是这部分“漏网之鱼”造成的。

[8]其他类型的一个例子是重力波，也称为g模（g来自“重力”的英文gravity）；其他周期的一个例子是所谓的“160分钟振荡”（现代观测并未证实这一周期，因此不排除该周期是以前观测留下的错误）。另外要提醒读者注意的是，日震学观测需要高超的数据分析能力，因为

我们观测到的振荡实际上是数以百万计的不同模式振荡相互叠加的结果，必须使用高超的观测技术及数学手段才能提取出有用数据。

[9] 第一拉格朗日点是太阳与地球连线上距地球约150万千米的一个点，该点上的卫星在太阳与地球引力的共同作用下，将会以与地球相同的公转周期绕太阳运动，从而始终保持在太阳与地球的连线上。

[10] 这种修正的一个著名例子是对对流区厚度的修正，标准模型原先采用的对流区厚度偏小。



绘画：张京

13 谜团锦簇的太阳大气层

我们的太阳故事到这里已接近尾声了。在本章中，我们将探索太阳的最后一个组成部分：大气层。那是一个肉眼通常难以窥视的地方，但利用各种仪器的帮助及日全食的机会，天文学家们已经对它进行了颇为细致的观测。观测的结果如何呢？概括地说是四个字：谜团锦簇。事实上，在这个谜团锦簇的太阳大气层中，我们将要面对的谜团可能要比在前面各章中遇到的加起来还多。这是因为太阳大气层比太阳内部更复杂吗？未必。更有可能的原因是我们对太阳大气层的观测远比对太阳内部来得细致。有一句西方俗语说得好：魔鬼存在于细节之中（The devil is in the details），太阳大气层无疑就是一个例子。

我们在第12章中已经提到过太阳大气层的某些部分。比如厚度约500千米的温度最低层，厚度约2 000千米的色球层等。本章的探索将从色球层开始，因为它将带给我们第一个谜团——色球层反常高温之谜。

色球层这一名称来自希腊文词根“chromos”，含义是颜色（color）。我们在第6章中介绍过的太阳元素“氦”就是在色球层的光谱中被发现的。色球层的一个很引人注目的特点，就是一反太阳内部温度自核心向外逐层递减的趋势，出现了外层温度比内层更高的奇怪现象，即所谓的色球层反常高温。这种反常高温的原因是什么呢？或者更具体地说，它所需的能量从何而来呢？科学家们进行了长期研究，并提出了一些可能的机制。目前看来比较靠谱的机制主要有两种：一种是声波加热，另一种是阿尔文波（Alfvén wave）加热。

声波加热顾名思义，就是认为色球层反常高温所需的能量来自于声

波。太阳上的声波我们在第12章中曾经介绍过，它们以太阳的肚子为谐振腔，形成很多模式，并且如第12章的注释所说，在被对流区顶部反射时会有一部分“漏网”，形成诸如“五分钟振荡”那样的大气层振荡现象。那么，那部分“漏网之鱼”的最终命运会如何呢？答案是：很悲惨。一般认为，色球层就是它们的葬身之地，而它们葬身之时留下的唯一“遗产”就是能量——维持色球层反常高温所需的能量。这就是所谓的声波加热机制。不过这种机制有一个显著的缺点，那就是只在能量需求不大的色球层下部才比较有效，在真正急需能量的色球层上部却不够“给力”。计算表明，大部分声波根本没到色球层上部就“出师未捷身先死”了。

那么色球层上部高温所需的能量又来自何方呢？一般认为，也是来自一种波，叫做“阿尔文波”（Alfvén wave）。那是一种沿着下文将要介绍的太阳磁场中的所谓磁通量管（magnetic flux tube）传播的波，是瑞典物理学家阿尔文（Hannes Alfvén, 1908—1995）提出的。研究表明，这种阿尔文波可以远比声波传得更高，从而可以为色球层的上部送去“温暖”，这就是所谓的阿尔文波加热机制。不过，声波加热与阿尔文波加热这两种机制的“高低搭配”是否算是解决了色球层反常高温之谜呢？目前还没人能打保票，因为很多细节仍有待完善。

反常高温并不是色球层带给我们的唯一谜团，色球层中的另一种现象也对天文学家们提出了挑战，那就是所谓的针状物（spicule）早在1877年，意大利天文学家塞奇（Angelo Secchi, 1818—1878）就在色球层的边缘上发现了这种形如尖针的东西。后来的观测证实，那是一些底面积与四川省差不多大，高度相当于几百个珠穆朗玛峰的“小火舌”（图13.1）。这样的“小火舌”在整个色球层中通常有六七万个之多，它们的温度在1万度左右，底部物质以每秒几十千米的速度向上升腾。这些“小

火舌”的成因也是一个谜团，一般认为，它们与“五分钟振荡”有着密切关系，但很多细节同样有待完善。

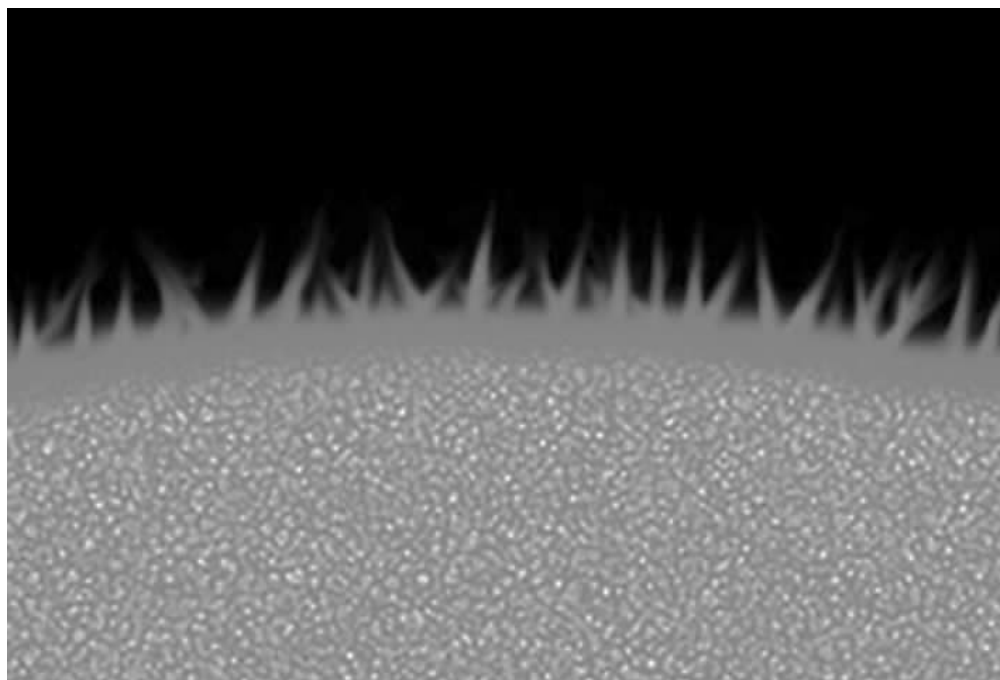


图13.1色球层中的“小火舌”——针状物

不过，我们对色球层的理解虽然是“缺点错误在所难免”，总体来说还算差强人意（这个词常被误用为“不能使人满意”，这次是按正确的含义来用，表示“勉强使人满意”）。与之相比，色球层以上部分的麻烦可就大多了。经过一个厚度200千米左右的“薄如蝉翼”的过渡区（transition region），太阳大气层的温度剧增到了100万度以上。那里便是日全食相片中最吸引眼球的东西——日冕（corona）。日冕这一名称的希腊文词根表示的是王冠（crown）。我们小时候画太阳时，通常会画上一些象征阳光的锯齿状轮廓，日冕就有点像那轮廓，只是形状要复杂得多。日冕不仅形状复杂，其他方面也很复杂，比如它的密度分布极不均匀，有些地方甚至有巨大的空洞，称为“冕洞”（coronal hole）；它的温度分布也同样极不均匀，平均为100万～300万度，局部的低温和高温则比比皆

是。^[1]

日冕给天文学家们带来的谜团是显而易见的，那就是它为什么会有如此高的温度？这个谜团与色球层的反常高温之谜如出一辙，“难度系数”却高得多。因为声波与阿尔文波这两种为色球层送去“温暖”的加热机制，对于“高高在上”的日冕来说，都显得有些“力不从心”。这其中声波加热机制连色球层上部都难以到达，自然就甭提了。阿尔文波的情况要强一些，它能为日冕提供一部分能量，但数量却不够。因此，解决日冕高温之谜还需要有其他机制的加盟，这其中比较有希望的一种机制，是美国天体物理学家帕克（Eugene Parker, 1927—）在1988年提出的所谓“纤耀斑”（nanoflare）加热机制。按照这种机制，日冕中常常会发生微型爆炸，即所谓的“纤耀斑”，这是我们将在下文中介绍的耀斑

（flare）的“迷你版”，威力“仅”相当于几百个氢弹，次数却很频繁，足以为日冕送去大量“温暖”，甚至有可能是日冕能量的主要来源。当然，这后一点目前谁也说不准。一般认为，太阳上不同规模耀斑的数量分布与自然界的很多其他随机现象一样，近似地满足所谓的幂律分布（即数量与规模之间满足近似的幂函数关系）。按照这种分布，“纤耀斑”的存在是不成问题的，数量也应该是比较大的，但是否大到能为日冕提供足够能量的程度，则与幂律的幂值有关，人们对之尚有争议。纤耀斑加热机制如此，其他机制的情况也大致相若，各有各的争议之处。那些争议不解决，日冕的高温之谜当然也就无法解决。

日冕不仅温度极高，体积也极为庞大。看过日冕相片的读者想必都对日冕的体积留有深刻印象。从相片上看，日冕有时能延伸到几个太阳半径处，其体积之大甚至超过了太阳本身。与地球大气层（厚度约100千米）的体积仅为地球体积的百分之几相比，日冕的体积无论绝对值还是相对值都是极为巨大的。但这是否就是太阳大气层的边缘呢？答案是

否定的。事实上，早在20世纪50年代中期，英国天文学家查普曼（Sydney Chapman, 1888—1970）就曾提出，日冕虽然看上去只延伸到几个太阳半径处，但如此高温的气体是不可能被禁锢在那样“小”的范围之内的。据他估计，日冕应该能一直延伸到地球轨道以外。（感兴趣的读者请思考一下，这与地球没有被“烤焦”这一显而易见的经验事实是否矛盾？）

与查普曼这一看法的提出几乎同时，德国天文学家比尔曼（Ludwig Biermann, 1907—1986）也提出了一个与太阳外围有关的想法，那就是我们如今称之为太阳风（solar wind）的想法。比尔曼的想法源于一个众所周知的现象，即彗星的尾巴——彗尾（comet tail）——总是沿背离太阳的方向伸展的，哪怕当彗星本身背离太阳运动时也是如此，仿佛是被一股来自太阳的“风”吹起来的。^[2]

查普曼与比尔曼的想法看起来很不相同，在空间范围上却存在巨大的重叠，它们彼此间是否有关联呢？上文提到过的美国天体物理学家帕克作出了肯定的回答。他的研究表明日冕不可能是一团宏观上静止的弥散气体，而应该是一种向外运动的“气流”——即“太阳风”。这样他就把查普曼的巨型日冕与比尔曼的太阳风联系了起来，并且他为太阳风构筑了一个具体的模型。有意思的是，帕克这项如今被视为太阳风研究的开山之作的重要研究，在当年却遭到了审稿人的拒绝，后来亏得著名美籍印度裔天体物理学家钱德拉塞卡（Subrahmanyan Chandrasekhar, 1910—1995）的慧眼相识，才得以发表。

有关太阳风的猜测很快就得到了观测证实。自20世纪50年代末起，苏联的“月球”（Lunar）系列航天器、“金星一号”（Venera）行星探测器，以及美国的“水手二号”（Mariner 2）行星探测器都先后观测到了太

阳风。^[3]如今我们知道，太阳风是太阳大气层外围的一种相当显著的现象，每秒钟从太阳上带走将近200万吨的物质。当然，这一乍看起来很惊人的数量相对于太阳的巨大质量来说只是“毛毛雨”（感兴趣的读者可以利用本书提供的数据，估算一下这种“毛毛雨”可以“下”多少年）。太阳风除了造成彗尾的“轻舞飞扬”外，更显著的结果是在太阳周围的星际介质中“吹泡泡”，被它吹出的泡泡称为“日球层”（heliosphere），其边缘通常被广义地视为是太阳大气层的边缘。那么这个边缘究竟在哪里呢？一般认为，是在离太阳100~200天文单位（即150亿~300亿千米）处。这个距离远远超过了所有太阳系已知行星的轨道半径，因此从某种意义上讲，我们一直都生活在太阳的大气层中——不仅我们如此，所有已知的太阳系行星都是如此，真所谓“普天之下，莫非王土”。^[4]

作为谜团锦簇的太阳大气层的成员，色球层和日冕都给天文学家们出了难题，太阳风也不例外。迄今为止，人们尚未找到一个能全面描述太阳风特征的模型。观测表明，太阳风的“风速”约在每秒200~800千米之间。其中“风速”在每秒400千米以下的被称为慢太阳风（slow solar wind），“风速”在每秒400千米以上的则被称为快太阳风（fast solar wind）。太阳风模型必须解释，却迄今尚不能给出满意解释的一个老大难问题，就是这两者——尤其是快太阳风——的起源。一般认为，快太阳风粒子的起源需要一个合适的加速机制，这种机制与日冕的加热机制很可能存在密切关系，甚至有可能是同源的（比如阿尔文波有可能对两者都起着重要作用），但具体如何，还有待进一步研究。

以上就是对太阳大气层各主要成员的大致罗列。小结一下的话，那么太阳大气层底部有一个很“凉快”薄层叫做温度最低层，自那以上温度不降反升，在温度不算太高的色球层中，有几万个底面积“只有”四川省那么大，高度“只有”珠穆朗玛峰的几百倍那么高的“小火舌”。在那以

上，经过一个“薄如蝉翼”的过渡区，是体积大得惊人的日冕，那里的物质极度稀薄，温度却高得邪乎，还时不时地有几百个氢弹炸来炸去。再往外，则是以比火箭还快几十倍的速度劲吹到几百亿千米外的太阳风。这就是太阳大气层的日常景观，这种景观有一个很文雅的名称，叫做“宁静太阳”（quiet sun）——别揉眼睛，您没看错，是叫做“宁静太阳”。

“宁静”的意思就是没什么“活动”。“宁静”尚且如此，那么有“活动”时又会怎样呢？

太阳大气层中的各种“活动”有一个很没创意的共同名字，就叫“太阳活动”（我们在第12章末尾已经提到过这一名称）。在太阳活动中，资格最老的是黑子（彩图12）。早在两千多年前，中国古籍中就出现了有关黑子的记载，西方人记载黑子也有一千多年的历史。不过黑子的发现虽早，长期以来却很少有人认清它的真面目，而常常把它当成所谓的行星凌日（planetary transit），即行星从太阳前面经过。直到17世纪初，才由伽利略明确提出黑子是太阳自身的“污点”。这个看似稀松平常的结论，对当时的“天贵地贱”观念给予了当头一击（参阅第3章）。^[5]不仅如此，这一结论还给另外一些人造成了困扰，比如我们在第8章中提到过的英国天文学家赫歇耳，就被黑子的“黑色”所误导，而以为太阳表面很凉快，甚至有可能存在生物。

那么这个蒙蔽了大伙儿几百年，甚至把赫歇耳那样的著名前辈都给误导了的黑子到底是什么东西？它为什么会出现在光芒万丈的太阳上呢？这些谜团直到20世纪初才逐渐揭开面纱。首先被搞清楚的是，黑子展现的绝不是赫歇耳所幻想的有可能有生物栖息的凉快表面，它的温度实际上高达4 000度左右（请读者想一想，我们介绍过的哪一种手段可

以使我们得到这一信息），足以熔化素以耐高温著称的金属元素钨。黑子之所以呈现黑色，完全是因为它不幸生在了一个比它更“白”的环境里（黑子周围太阳物质的温度比它高了一千多度），从而被衬黑了。如果我们能将黑子周围的阳光全部滤掉的话，黑子非但不黑，反而会呈现出耀眼的光芒。另外，黑子看上去很小（黑子的英文名字中的“spot”和中文名字中的“子”都有“小”的意思），那也只是相对于太阳的庞大而言的，它的实际块头可不小，即便小黑子的面积也有新疆那么大，大黑子更是可以吞下整个地球。

黑子的“黑”和“小”这两个品性算是被搞清楚了，但它为什么能“出高温而不染”，维持这么“低”的温度呢？这可就不是一个简单问题了。它所牵扯到的是太阳大气层中最重要的幕后推手：太阳磁场。

1908年，美国天文学家黑尔（George Hale, 1868—1938）通过光谱学手段发现，黑子中存在强度约为0.3特斯拉（0.3 T）的磁场。^[6]这是很强的磁场，比地球磁场强一万倍左右，也比太阳表面的平均磁场强得多。这一强磁场的存在为揭开黑子的“维稳”之谜提供了线索。这线索就是：“维稳”离不开“警力”，而磁场恰恰可以充当这种“警力”，因为它会产生一种特殊的压强，叫做磁场压。正是在磁场压的帮助下，温度较低（从而内部压强也较低）的黑子才能抗衡住外部物质的高压而维持稳定（黑子一般能维持几天至几星期，在太阳大气层中算是比较稳定的结构）。进一步的计算还表明，强磁场对太阳物质的对流会产生抑制作用，使得能量无法经由对流有效地传至黑子内部，这就解释了黑子温度偏低的原因。更令人欣慰的是，在理论上还可以证明，太阳物质的运动会通过一种所谓的“对流坍塌”（convective collapse）现象，而自发地造成某些区域的磁场强度增加。这种磁场强度增加的区域通常呈管状分布，被称为“磁通量管”。磁通量管伸出或插回太阳表面的地方通常就会

形成黑子。由此可见，磁场的存在除了能解释黑子的主要特征及“维稳”机制外，还可以解释它的成因。不仅如此，由于磁通量管的伸出与插回总是“成双成对”，且极性相反的，它还可以解释有关黑子整体分布的一个重要特征，那就是黑子常常成对出现，且每对黑子的磁场极性彼此相反。

但是，磁场的存在虽然解释了黑子的成因、主要特征及“维稳”机制，却不等于说黑子就没有带给我们其他谜团了，那样的谜团其实还不少。细致的观测表明，黑子有着复杂的结构，除了被称为“本影”（umbra）的黑色中央区域外，半径在5 000千米以上的大黑子四周通常还有所谓的“半影”（penumbra），它们的颜色较浅，包含了很多纤维状结构。更仔细的观测还表明，本影中有时会出现亮点，半影内则有时会出现旋涡状结构，在黑子消亡前，本影内有时还会出现明亮的桥状结构。这些细致结构的成因及演化目前都还是有待探索的谜团。

太阳活动的另一个重要成员叫做耀斑（彩图13）。如果说黑子是一种很低调的太阳活动，那么耀斑就恰好相反，它极为张扬，是“爆炸式”的活动——这可不是比喻，因为它实际上就是爆炸。一个大耀斑通常可以释放出几十亿亿焦耳的能量，相当于一百亿个百万吨级氢弹同时爆炸！在大耀斑爆发时，太阳大气层的局部温度可以在短时间内升高到2 000万~3 000万度，比太阳核心的温度还高。耀斑的威力极为惊人，虽然发生在一亿五千万千米之外的太阳上，却足以对地球产生显著影响。事实上，1859年9月，人类记录下的第一个大耀斑就是以它对地球的显著影响而引起人们的注意的，它所发射的带电粒子流猛烈撞击地球磁场，产生的极光（aurora）一直延伸到赤道附近，使无数没有机会前往极地的人领略了一次天象奇观。随着技术的发展，耀斑对地球的影响得到了越来越多的显现机会。1984年4月，一个普通耀斑中断了美国

总统里根（Ronald Reagan，1911—2004）的“空军一号”专机与地面的通信，使美国情报部门大为紧张，以为是苏联人在做手脚。2006年12月，一个小耀斑将正在进行舱外作业的国际空间站的宇航员逼入“发现者”号航天飞机仓皇躲避（他们实际上还是慢了一步，所幸那只是小耀斑，他们所受的辐射剂量不大）。

耀斑给我们提出的问题是显而易见的：那就是它为什么会发生？它的能量来自何方？科学家们对这些问题进行了长期研究。与黑子一样，一般认为耀斑的发生也跟太阳磁场有着密切关系。一种目前比较流行的观点认为，耀斑的发生很可能是磁通量管的重组造成的。太阳上的磁通量管就像橡皮筋一样，可以通过应力等形式储存能量（上文提到的阿尔文波之所以能沿磁通量管传播，也正是因为这一特性）。随着太阳物质永不停息的运动，磁通量管会被拉伸、扭曲、缠绕，这些过程会大大增加储存在磁通量管中的能量。但这种过程是不可能永远持续下去的，因为能量越高的状态就越不稳定，最终将会有有一个时刻，如同橡皮筋突然断裂一样，磁通量管发生重组，由拉伸、扭曲、缠绕后的复杂状态一举重组回相对简单的状态，并将因拉伸、扭曲、缠绕而储存起来的能量在很短的时间内释放出来，由此形成的就是耀斑。这个过程有点像小孩搭积木，随着积木越搭越高，势能越积越多，它会变得越来越不稳定，最终会有有一个时刻，积木突然垮塌，原先储存在积木中的势能被突然释放出来（释放的形式之一就是撞击桌面或地面发出的轰然之声）。

这种机制得到了一定程度的观测支持。人们发现，耀斑通常发生在极性变化较为复杂的黑子区域附近，那里不仅有强磁场，而且磁通量管的分布较为复杂，容易发生上述机制所要求的拉伸、扭曲、缠绕等过程。但这种机制也并非没有问题。比如前面提到的帕克在20世纪60年代曾对磁通量管的重组进行过估算，结果发现那需要几小时甚至几天的时

间才能完成，而耀斑的能量释放过程往往持续不到一分钟，两者大相径庭。为了解决这一问题，科学家们提出了许多修正方案，比如有人提出磁通量管的重组只需发生在一小段而不是整段上，从而能在短得多的时间内完成。也有人提出磁通量管在重组前会破碎成许多小圈，从而大大增加接触面积，使重组得以“提速”，等等。总之，方案人人会提，各有不同巧妙，但越是巧妙的方案所需的观测证实通常也越精密，其中很多都超出了目前的观测能力，因此耀斑的发生机制到目前为止还是一个谜，探索的道路还很漫长。

除耀斑外，太阳大气层中的另一种爆发现象近来也受到了越来越多的关注，那就是所谓的日冕物质抛射（coronal mass ejection）。一定规模的日冕物质抛射的发生频率从几天一次到一天几次不等，它们与耀斑的主要差别——顾名思义——是会抛射出大量的“物质”，主要是带电粒子，数量约有几十亿至上百亿吨，飞行速度约为每秒几百千米，所携带的动能与大耀斑的总能量相近。[\[4\]](#)这些质量相当于几万艘万吨巨轮的带电物质若不幸与地球相遇，将会造成大型的地磁暴，其威力往往要比耀斑造成的地磁暴更厉害——这也正是人们越来越关注日冕物质抛射的主要原因。

1989年3月，一次小规模日冕物质抛射引发的地磁暴，造成了加拿大魁北克省电力系统的崩溃，并使得极光范围一直延伸到美国南部的得克萨斯州，很多人甚至担心美苏双方的军事通信系统会因地磁暴的干扰而触发核大战。小规模日冕物质抛射尚且如此厉害，大规模日冕物质抛射倘若击中地球，结果更是不堪设想。2009年5月，美国国家海洋与大气管理局（NOAA）发布了一份太阳风暴（solar storm）警报，认为2013年有可能会是太阳活动较频繁的年份，其中某些太阳活动的规模有可能不在1859年那次大耀斑之下。这其中最令人担忧的就是大规模日冕

物质抛射击中地球的可能性。有人也许会说：1859年的大耀斑不就是一次天象奇观吗？并没有造成严重后果啊？是的，1859年的大耀斑确实没有造成严重后果，但今天的人类社会早已不是1859年那样子了。我们今天引以为傲的现代文明已经极大地依赖于各种各样的电子及电力设备。在这种情况下若被大规模日冕物质抛射所击中，后果将远不是天象奇观那么浪漫，而很可能是灾难性的，比如很可能出现全球性的电力中断（图13.2）。现代社会是如此彻底地依赖电力，就连像加油站那样以提供其他能源为目的的系统，也已经离不开电力。一旦发生全球性的电力中断，后果将不堪设想，那时想要修复哪怕一座爆炸了的变压器都会极度困难。科学家们估计，那样的灾难倘若发生，人类恐怕要用几年的时间才能逐渐恢复，在此期间，连食物和水的供应都会出现困难，其结果将是比任何大型战争或重大瘟疫还要可怕得多的生命损失，就连发达国家也难以幸免（事实上，在那样的灾难面前，发达国家反而有可能更脆弱，因为它们比其他国家更依赖于电子及电力设备）。



图13.2 太阳风暴假想图

当然，这只是最坏情形，即便真的发生大规模日冕物质抛射，它直接击中地球的概率也并不大，但想到人类只能仰仗大自然的概率恩赐来侥幸度日，毕竟是一种很无奈的感觉。太阳对于我们实在太重要了，为今后有可能发生的来自于太阳的灾难预做准备，是除好奇心之外，人类研究太阳的一个最重要的动机。事实上，最近几年来，人们已经开始建立所谓的“太空气象”（space weather）系统，试图对主要由太阳活动引起的地球附近的空间环境变化进行研究及预测，以服务于我们这个脆弱星球上的脆弱系统。但到目前为止此类系统尚不能作出真正意义上的预测（比如前面提到的美国国家海洋与大气管理局有关2013年太阳风暴的警报不仅在时间上是粗略的，也规模上也完全有可能是错误的）。它所能做的真正可靠的事情，只是利用带电粒子从太阳运动到地球所需的几天时间，通过已经发生的太阳活动，来为我们提供几天的预警时间而已。

那么，日冕物质抛射是如何发生的呢？很不幸，这也是一个未解之谜。人们提出了很多理论，但各有各的问题。比如很多人曾经认为，日冕物质抛射与耀斑有着直接关联，甚至完全就是由耀斑造成的。但后来的研究发现，很多大耀斑并不导致日冕物质抛射，而某些小规模の日冕物质抛射则似乎没有与之相伴的耀斑。因此日冕物质抛射与耀斑之间看来并不存在可靠的因果关联。但一般认为，两者之间虽没有因果关联，却也绝非毫不相干。佐证这一点的最好证据，就是几乎所有大规模の日冕物质抛射都有与之相伴的耀斑。另外，日冕物质抛射与耀斑一样，都几乎铁定是与太阳磁场存在极密切的关系，而且极有可能都是磁通量管的重组造成的（事实上，日冕物质抛射很可能是整段磁通量管遭到“抛弃”所致）。只不过究竟什么样的磁通量管分布会造成耀斑，什么样的磁通量管分布会造成日冕物质抛射，目前还不得而知。

从黑子到耀斑，再到日冕物质抛射，太阳活动的形式是丰富多彩的。那么，在这些活动中有没有什么共同特征呢？有，那就是太阳周期（solar cycle，图13.3），那也是我们在结束走马观花般的本章前，最后要介绍的话题。

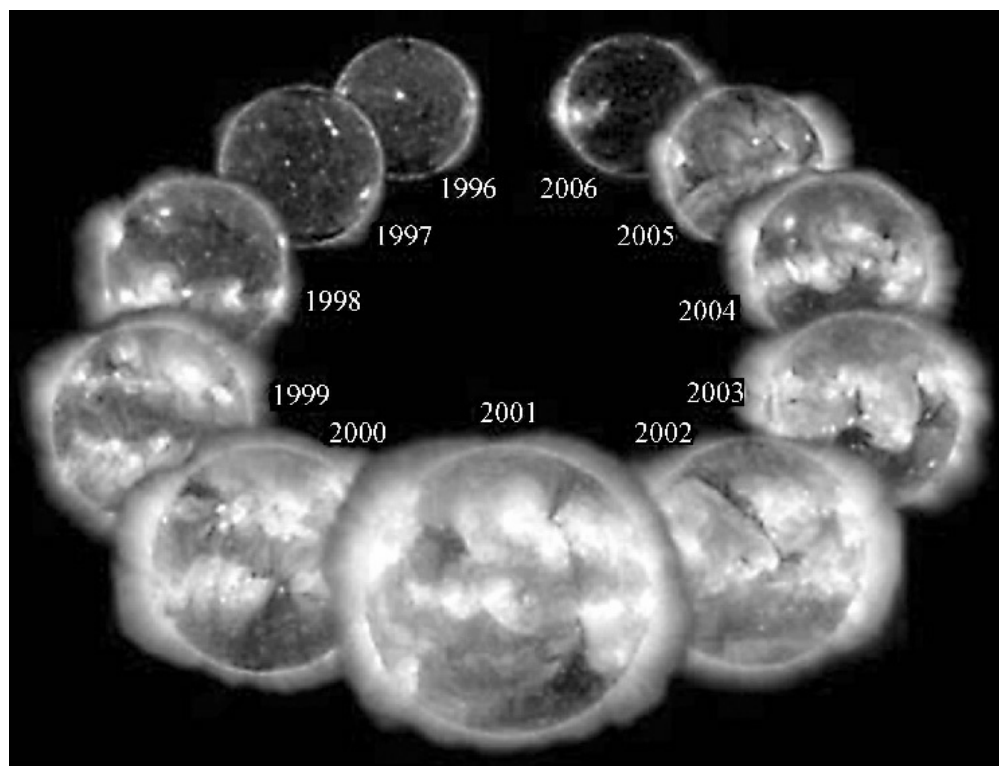


图13.3 太阳周期（顶端为极小年，底部为极大年）

太阳周期是德国天文学家施瓦布（Heinrich Schwabe, 1789—1875）通过对太阳黑子数目长达17年的持续观测，并借鉴历史数据所发现的。它最初的含义是黑子数目的变化周期。但后来的研究发现它同时也影响着耀斑、日冕物质抛射等其他太阳活动，甚至对太阳的光度及太阳风的强弱也有一定影响，可以说是所有太阳活动的共同特征。另一方面，在揭示出太阳周期所具有的广泛影响力的同时，它作为周期现象的品质却在下降。现代观测表明，太阳周期并不是严格意义下的周期现象，它的平均值约为11年^[8]，但有时可以短至9年，有时又可以长达16

年，甚至还出现过长达几十年没有显著活动的所谓“蒙德极小期”（Maunder Minimum）。不过，尽管有这样的起伏，太阳活动的规律性仍是足够显著的，起码存在着近似意义上的周期性，而绝不是随机现象。在一个太阳周期中，太阳活动最少的年份称为太阳活动极小年（solar minimum），太阳活动最多的年份则称为太阳活动极大年（solar maximum）。前面提到的美国国家海洋与大气管理局有关太阳风暴的警告，其实也是对太阳活动极大年的预测，即认为2013年很可能是即将来临的一个太阳活动极大年。[\[9\]](#)

尽管太阳周期不是严格意义下的周期，但作为一种显著的近似现象，它的存在依然是需要解释的。为了寻找这种解释，科学家们付出几十年的艰辛努力，可惜的是——读者们一定猜到我要说什么了——迄今为止这也仍是一个未解之谜。

当然，这并不是说几十年的努力毫无成果。相反，成果不仅有，而且还不小。在这里我们介绍其中较受青睐的一个，即所谓的“发电机机制”（dynamo mechanism）。在介绍之前，首先让我们一同分析一下，解释太阳周期的切入点应该在哪里？大家请不要拍砖，这并不是拿一个研究性的问题来卖弄知识或为难读者。事实上，对太阳周期的具体解释虽然是极困难的研究课题，切入点的选择却是相当显而易见的。通过本章的介绍，我们已经知道太阳活动有一个共同的幕后推手，那就是太阳磁场。而我们前面刚刚介绍过，太阳周期是所有太阳活动的共同特征。既然一组现象有一个共同的幕后推手，那么它们的共同特征最有可能归因于什么呢？当然就是那个共同推手，即太阳磁场。因此，解释太阳周期的切入点应该是太阳磁场，这也正是发电机机制的切入点。发电机机制的精髓之处，就是将太阳磁场本身的演化也纳入了解释范围之内，从而在很大程度上具有统领整个领域的潜力。

发电机机制出现于20世纪60年代，此后经历了几次起落。它的早期设想是这样的：在太阳活动的极小年中，太阳内部存在着弱磁场（像太阳这样的巨型等离子气团中存在弱磁场并不是难以想象的事情，不存在反倒是难以想象的）。这种场在太阳内部主要由南极指向北极，被称为极向场（poloidal field）。极向场在太阳内部自转不均匀性的带动下，将会发生扭曲，使磁通量管沿赤道方向遭到拉伸，形成所谓的环形场（toroidal field）。随着扭曲的加剧，环形场的强度会持续增强。计算表明，当环形场的强度增加到一定程度时，磁通量管将会出现上浮的现象。不仅如此，磁通量管在上浮过程中还会产生新的极向场^[10]，从而构成极向场与环形场之间的相互支撑。当磁通量管浮出太阳表面时，将如上文说过的那样，在伸出和插回太阳表面的地方形成一对磁场极性相反的黑子。粗略的分析表明，这种机制有一个很漂亮的地方，那就是可以解释有关黑子分布的所谓斯波勒定律（Spörer's law），即在每个太阳周期中，自极小年到极大年，太阳黑子几乎总是首先出现在高纬度区域，然后向低纬度区域扩张，而且南北半球上的黑子对有着相反的极性。那么太阳周期又是如何出现的呢？是因为磁通量管的相互湮灭。当南北半球那些极性相反的黑子对扩张到赤道附近时，磁场方向相反的磁通量管会相互湮灭，从而使磁场分布逐渐回到最初时的弱磁场状态——这就是一个新周期的开始。

这些结果初看起来很令人振奋，不仅对太阳周期作出了说明，还附送了对斯波勒定律的解释，简直是“买一送一”。可惜好景不长，问题很快就出现了，而且恰恰就出在那附送的部分上。原来，发电机机制的早期研究有一个先天不足，那就是对太阳内部的自转状况一无所知。当然，在那种早期研究中，无知有无知的快乐，那就是便于作假设，因此当时人们对太阳内部的自转状况作了相当任意的假设，目的之一就是解释斯波勒定律。但不久之后，我们第12章所介绍的日震学手段就粉墨登

场了，它无情地粉碎了那些无知年代的快乐假设，因为它所探测到的太阳内部的实际自转状态与早期所假设的并不一致。那么，利用太阳内部的实际自转状态是不是仍能解释斯波勒定律呢？很遗憾，不能。事实上，它很不幸地会导致黑子分布规律的逆转，即变成了从低纬度向高纬度扩张，与斯波勒定律恰好相反。这样一来，原先“买一送一”的友情馈赠反倒变成了躲都躲不开的“硬性搭售”。更麻烦的是，早期发电机机制被认为是在对流区中起作用的，新的研究却发现，太阳对流区的环境过于恶劣，到处是强劲的湍流。在那样的环境下，磁通量管很快就会被撕碎，从而根本就不可能有时来孕育足够强的磁场。这样一来，发电机机制连对流区这一“老巢”都失守了。

这些问题给发电机机制蒙上了巨大的阴影。但天文学家们没有放弃，而是积极寻找着解决之道。到了20世纪90年代中期，一种新设想为发电机机制注入了新活力。这种新设想就是“迁居”——将发电机机制的作用地点由对流区迁移到差旋层中。差旋层我们在第12章末尾曾经提到过，它是对流区底部以下的一个薄层，是太阳刚性自转与非刚性自转的交界层。这样的交界层自然也会造成极向场的扭曲，从而也可以启动发电机机制^[11]。但与对流区不同的是，差旋层由于地处对流区底部以下，因而是一个相对稳定的地方，发电机机制可以在那里从容不迫地积累强磁场。不仅如此，计算表明，这种新设想所导致的黑子分布规律与斯波勒定律又重新一致了，从而排除了一个重大硬伤，使局势进一步明朗起来。不过硬伤虽已被排除，软伤的有无却是谁也无法打包票。发电机机制作为一个非线性模型，存在着诸如混沌之类的复杂性，而太阳周期本身也是一种复杂现象，这两者的拟合在细节上几乎无可避免地存在很大的不确定性，使得谁也无法保证目前这明朗局势是否是昙花一现，甚至只是回光返照。我们只能说发电机机制是迄今为止有关太阳磁场及太阳周期的比较有希望的机制。

我们对太阳各部分的介绍到这里就全部结束了，这些介绍归根到底是从空间角度加以介绍。也许让读者不无失望的是，我们居然是在一堆谜团的重重包围之下结束了漫长的旅程。但换个角度想想，这岂非恰恰是这一领域充满生机的明证吗？在第14章，同时也是本书的最后一章中，我们将由空间转往时间，去进行一次时间之旅，探索太阳的过去和未来。

[1] 日冕的温度是用光谱学手段测定的。我们在第6章中提到过，19世纪下半叶，人们曾在日冕中发现过许多新谱线，甚至一度以为发现了新元素，后来才意识到它们来自于高度电离的金属离子。这一点虽然打碎了发现新元素的美梦，却给了人们一个估算日冕温度的机会，因为电离需要高温，要想产生那些日冕谱线所对应的高度电离的金属离子，必须有100万~300万度的高温，这就是日冕温度的观测由来。

[2] 看过彗发相片的读者可能会觉得奇怪，彗发“仿佛是被一股来自太阳的‘风’吹起来的”是一种很容易萌发的感觉，为何直到20世纪50年代才由比尔曼提出太阳风的想法呢？这是因为对这种感觉有一种更简单的解释，那就是彗发是被阳光本身的光压“吹”起来的。比尔曼是在排除了这种可能性之后才提出“太阳风”的想法的，而非单纯猜测。另外还要提到的是，比比尔曼早一个世纪的19世纪中叶就有人提出过类似于太阳风的想法，起因是太阳耀斑与地磁暴之间存在时间关联，仿佛后者是由前者所发射的粒子流引起的。（请读者想一想，这种“太阳风”与比尔曼的有何区别？）

[3] 太阳风的存在首先由航天器而非地面观测站所证实不是偶然的，因为太阳风是一种以电子和质子为主的带电粒子流，它们会被地磁场所排斥，从而很难被地面观测站所直接探测。另外要提到的是，太阳

风早已不再被视为是日冕的一部分了，甚至连它是否属于太阳大气层，有些文献也有不同的选择。

[4] 日球层的边缘不是一个单一概念，它自内向外包含了三个层面：太阳风的风速由超声速变为亚声速的地方被称为“终端激波”（termination shock）；太阳风的压强减弱到无法再推动星际尘埃（即两者压强相等）的地方被称为“日球层顶”（heliopause）；星际介质的运动首度因遭遇太阳风而减速的地方被称为“弓形激波”（bow shock）。另外要提到的是，有些人将日球层的边缘或其中某一个层面——比如“日球层顶”——视为太阳系的边界，那是不合适的，因为它无法涵盖被认为是太阳系最外围的所谓奥尔特云（Oort cloud）。

[5] 令人感慨的是，太阳黑子不仅在相对蒙昧的伽利略时代对某些观念造成了冲击，甚至在三百多年后的中国“文革”期间，还因为太阳所承载的政治象征而成为了禁忌，有些教师甚至因为讲授太阳黑子而被打倒，乃至打死。

[6] 黑尔的这一研究利用了荷兰物理学家塞曼（Pieter Zeeman, 1865—1943）在1897年所发现的所谓塞曼效应（Zeeman effect），它表明光谱线在磁场作用下会分裂，且裂距与磁场强度成正比，这一效应开启了用光谱学手段测算磁场强度的大门。

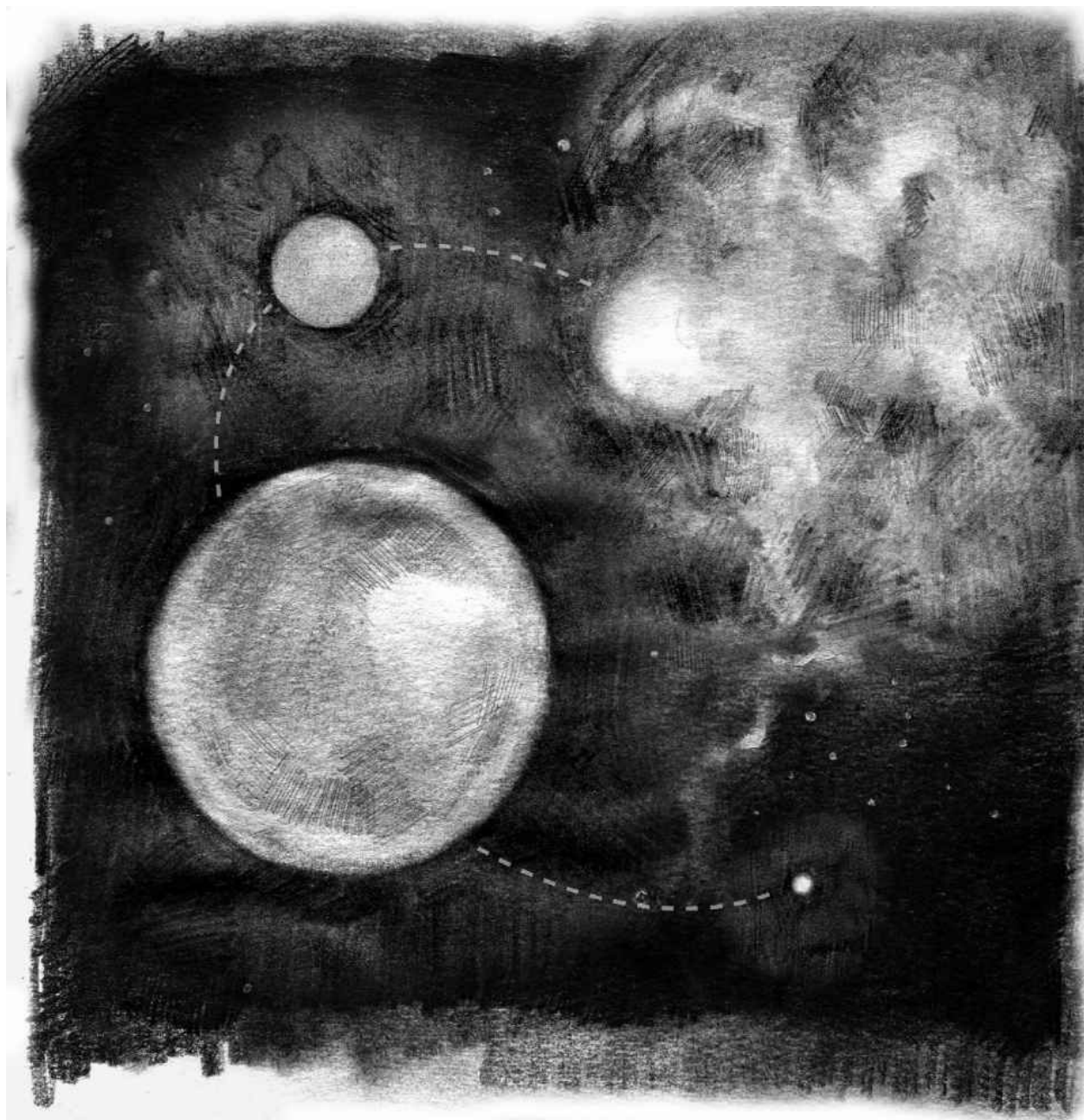
[7] 耀斑也能抛射物质，但数量要少得多。另外顺便说一下，无论耀斑还是日冕物质抛射，它们释放的能量虽然都极其可观（瞬间释放的能量能占到整个太阳能量的一个很可观的比例），但由于频谱分布极为宽广，在可见光范围内的相对比例很小，因此不是肉眼所能感知的。

[8]有一个细节在这里要提一下，那就是两个相邻的太阳周期在太阳活动的规模、分布等方面虽然相似，太阳磁场的极性却是相反的。如果把这个差别也考虑进去，那么太阳周期更确切地说是约为22年。这一点是上文提到的美国天文学家黑尔所提出的。

[9]如今回过头来看，在2013年所属的这次太阳周期中，太阳活动并没有如预期那样频繁，也并未对地球造成重大影响，因此美国国家海洋与大气管理局的警报在规模上是错误的。另外，标志这次太阳周期顶峰的黑子数目的最大值出现在2014年，与2013年作为太阳活动极大年的预期也有一定偏差。不过此类预测本就是粗略的，出现这种情形也并不意外。[2015年6月17日补注]

[10]这一点是美国天体物理学家帕克的贡献。他提出，在磁通量管的上浮过程中，因太阳自转而产生的科氏力（Coriolis force，地球上台风中心附近的气流旋转方向主要就是受其影响而形成的）将会使它形成新的扭曲，那些新扭曲在湍流的帮助下将会合并成携带极向场的新磁通量管。

[11]将发电机机制的作用地点移到对流区底部以下的想法，实际上早在差旋层被发现之前的20世纪80年代初就有人提出过，不过细节性的研究直到20世纪90年代中期才出现。



绘画：张京

14 太阳的过去和未来

按我们目前对物理学定律的了解，在四维的物理时空中，我们可以在三个空间维度中自由运动，却只能在时间中“随波逐流”，既不能回到过去，也无法随意前往未来。可人类的好奇心却偏偏对许多事物的过去和未来颇怀有浓厚兴趣，大至宏伟的宇宙，小至普通的生命，人类都想知道它的过去和未来。就连懵懵懂懂的小孩有时也会问：我是从哪里来的？我们太阳之旅的最后一站就要来回应这种好奇心，去探索太阳的过去（起源）是怎样的？它的未来（归宿）又将如何？这些问题不仅是我们的好奇心所系，而且与人类的命运不无关系——假如人类能在地球上存在得足够久的话。

但过去和未来既然是去不了的地方，我们又如何能探索呢？这个看似严重的问题并没有难倒科学家。事实上，我们这整个太阳故事，都是在探索去不了的地方：从太阳的核心到它的大气层，哪儿都不是我们能去的，但通过科学方法，我们依然得到了许多能与观测相印证，从而有很高可信度的结果。探索太阳的过去和未来也是如此。从某种意义上讲，它甚至比前面几章的探索还略微容易些，因为有大量的参照物可供借鉴。那参照物就是其他天体，它们各有各的年龄，有的尚处于“襁褓”阶段，有的是“小伙子”，有的已“人到中年”，有的则已步入“古稀”。通过对它们的观测，我们就能知道太阳的过去和未来大致会是什么样子的，并印证有关太阳演化的理论或假说。这就好比对不同年龄的人进行观察，我们就能知道自己的过去和未来大致会是什么样子的。

当然，我们也不能低估这种探索的难度。毕竟，判断天体的年龄是需要知识的，那种知识的获取则有一定的难度。事实上，单是破除天体

永恒完美的宗教式观念，从而使天体的年龄概念有意义，就经历了一个漫长的过程（参阅第3章）。对天体年龄的具体判断则更复杂，在理论和观测上都是很有难度的。直到今天，科学家们仍在为观测及判定某些特定年龄段的天体而努力着。这种难度导致的一个结果，就是有关太阳演化的早期理论大都带有较大的猜测性，而且主要局限于关注太阳的起源。至于太阳的归宿，在连太阳为什么会发光那样的问题都尚未得到解决的年代，是很难进行有价值的探索的，因此直到20世纪30年代之后，才有值得一提的发展。今天，随着太阳模型的逐步完善以及观测、计算技术的快速发展，我们对太阳演化的研究已变得越来越定量，也越来越有可信度了。当然，这一切距离“结题”无疑还差得很远，这么大的课题几乎注定是要长期探索的。

现代太阳演化理论的奠基者一般被认为是德国哲学家康德（Immanuel Kant, 1724—1804）。1755年，这位对自然科学怀有浓厚兴趣的哲学家在《自然通史和天体论》（*General History of Nature and Theory of the Heavens*）一书中提出了太阳系起源的所谓星云假说（nebula hypothesis）。康德这一假说的灵感由来是比较特别的，因为在他那个年代，人们对其他天体的年龄还缺乏认识，从而无法以它们为参照物。康德是从一个完全不同的角度，即太阳系与当时正被陆续发现中的星云之间的相似性，而得到的灵感。^[1]不过，康德毕竟是哲学家而不是科学家，他的著作虽然提出了星云假说的雏形，却也充斥着纯粹的臆想，比如他认为绝大多数行星上都存在居民，而且在这纯属臆想的东西之上，还言辞凿凿地提出了离太阳越远的行星上的居民越优秀那样的荒唐结论。^[2]对此，英国哲学家罗素（Bertrand Russell, 1872—1970）作过一句很幽默的评价：“这种见解作为地球人的谦虚值得赞许，但并无任何科学依据。”

继康德之后对星云假说作出重大贡献的是法国数学家拉普拉斯（Pierre-Simon Laplace, 1749—1827）。1796年，他在科普著作《宇宙体系论》（*The System of the World*）的附录中独立地提出了星云假说。拉普拉斯对星云假说的论述作为一部科普读物的附录，明显不如康德的论述来得详细，但却比康德的论述更紧凑、更物理、更注重于为假说寻找观测依据，并且也更注重于一些后来被证实为重要的因素，比如星云的自转。由于拉普拉斯在科学界的声望要明显高于康德，他对星云假说的“插足”在很大程度上促成了这一假说在即将到来的19世纪的流行。为了纪念拉普拉斯的贡献，后人将他的名字与康德并列，将星云假说称为康德-拉普拉斯星云假说，简称星云说。

虽然有拉普拉斯作后盾，星云说的“星运”依然不是一帆风顺的，在经历了19世纪的风光后，星云说开始面临越来越多的细节性困难，甚至一度走向了衰落。与这种衰落遥相呼应的，则是一些其他假说——比如灾变说（认为行星源于外来天体与太阳的碰撞）、爆发说（认为行星源于太阳的爆发）、俘获说（认为行星是太阳的“俘虏”）等——的“揭竿而起”。当然，即便在所谓的衰落期，星云说仍有一定的发展，只是不再像19世纪那样唯我独尊而已。20世纪80年代之后，随着新恒星周围的行星盘被观测天文学家所发现，星云说的命运迎来了新的转机。时至今日，星云说虽依然面临一些困难（太阳系作为一个比太阳更复杂的系统，它的起源理论存在困难是意料中的事），其主流地位却已基本无可撼动。而且星云说的困难大都集中在解释行星和卫星的某些特征上，对于太阳的起源，则几乎已达到了铁板钉钉的确切度。事实上，即便在所谓的衰落期，星云说的对手们也很少挑战它对太阳起源的描述，而将注意力集中在行星和卫星上。

有着两百多年历史的星云说奠定了研究太阳过去（起源）的理论框

架，那么对太阳未来（归宿）的研究又如何呢？这方面的研究历史要短得多。这是因为决定太阳未来的最重要因素：太阳的能量来源，以及那来源可以支撑多久，直到20世纪30年代末才得到初步解决（参阅第8章）。

下面我们就来介绍一下目前学术界有关太阳过去和未来的主流看法。

按照星云说，太阳（以及太阳系的其他主要天体）都是由所谓的星际云（interstellar cloud）坍缩而成的。但星际云有好几种类型，究竟什么样的星际云最有可能成为太阳或其他恒星的“摇篮”呢？要想回答这一问题，我们首先要对星际云的分类作一个简单介绍。星际云的分类是以宇宙中含量最丰富的元素——氢——的形态为依据的。具体地说，星际云通常分为三类，也称为三个区（之所以称为“区”，是因为不同类型的星际云所占据的往往是同一片大星际云的不同区域）：

（1）氢以离子状态存在的被称为HII区，它的温度最高（往往能达到10 000 K）。

（2）氢以原子形态存在的被称为HI区，它的温度适中（通常在100 K左右）。

（3）氢以分子形态存在的被称为分子云（molecular cloud，彩图14），它的温度最低（往往只有10 K左右，几乎是整个星系中最冷的物质），平均密度则较高（每立方厘米约有几百个粒子），相当于太阳附近星际介质密度的100倍以上，因而也被称为稠密云（dense cloud）[\[3\]](#)。

在这三类星际云中，最有可能成为太阳或其他恒星“摇篮”的是分子云。为什么呢？因为一片星际云要想坍缩，必须满足一个显而易见的条

件，那就是不能无限期地“维持现状”（否则坍缩就无从谈起了）。这一条件有一个学术名称，叫做不稳定性（*instability*）。那么，一片星际云怎样才会有不稳定性呢？答案是它的总质量必须足够大。而且星际云的温度越高、密度越低，出现不稳定性所需要的总质量就越大。这一结果是不难理解的，因为使星际云坍缩的动力是引力，负隅顽抗的则是它作为弥散气体所具有的压强。如果在两者的竞争中引力占到上风，并且在坍缩过程中能持续占到上风，坍缩就会很自然地发生，这样的星际云就有不稳定性。明白了这一点，也就不难理解“温度越高、密度越低，所需要的总质量就越大”了。因为温度高意味着压强大，密度低则意味着引力弱，两者都不利于引力。为了克服这些不利因素，引力必须倚仗它的“独门武功”，那就是总质量。因为总质量越大，引力就越强，这是压强所不具有的特点。因此在“温度越高、密度越低”那样的不利条件下，“所需要的总质量就越大”。

具体的计算表明，星际云产生不稳定性所需的最小质量正比于温度的 $3/2$ 次方，反比于密度的 $1/2$ 次方。这一结果是英国物理学家金斯（*James Jeans*, 1877—1946）在1902年给出的。为了纪念这一早期研究，人们把这种不稳定性称为金斯不稳定性（*Jeans instability*），相应的最小质量则称为金斯质量（*Jeans mass*）。

那么，分子云的金斯质量有多大呢？计算表明，约为太阳质量的几千倍（相应的线度约为几十至上百光年）。而HI区和HII区的星际云由于温度更高、密度更低，从而金斯质量更大，大到了实际上很少能达到的程度，因此不容易成为太阳或其他恒星的“摇篮”。^[4]

有了不稳定性，坍缩就将是不可能的，因为很多偶然因素——比如附近超新星爆发产生的激波，星际云的相互碰撞，星际云穿越星系旋

臂所遇到的干扰等——都能触发坍缩的发生。那么，坍缩的具体过程会是怎样的呢？质量约为太阳质量几千倍的分子云，会不会坍缩成一个质量达几千个太阳质量的超级恒星呢？答案是否定的。事实上，质量达几千个太阳质量的超级恒星不仅在观测上从未被发现过，在理论上也被认为是不可能的。这其中一个很重要的原因，就是在坍缩过程中，随着密度的增加，金斯质量将会减小（因为金斯质量反比于密度的 $1/2$ 次方），从而导致分子云中一些较小的部分发生独立的坍缩，使整片分子云分裂成小块。而且那样的小块随着密度的增加还会进一步分裂，使得最终形成的恒星质量远小于星际云的总质量。

但这种分裂也不会无限制地持续下去（否则就只有粉末而不会有恒星了）。当分子云的密度大到一定程度后，坍缩过程中释放出来的引力势能将被裹在其中而无法及时散去，这将造成分子云温度的上升，从而遏制金斯质量的进一步减小（因为金斯质量正比于温度的 $3/2$ 次方）。计算表明，这时候的金斯质量大约在太阳质量的几分之一至几倍之间，相应的分子云线度则在几千至几万天文单位之间。这种分子云被称为原恒星云（protostellar cloud），太阳系的前身就是那样一片原恒星云。

原恒星云通常是扁盘状的，这是因为星际云通常存在缓慢的自转，在坍缩过程中，这种看似不起眼的自转会使得平行于转轴方向的坍缩比垂直于转轴方向的坍缩更容易进行（因为后者会因角动量的守恒而旋转得越来越快，从而产生越来越显著的离心效应，阻碍坍缩的持续进行），由此导致的结果就是使分子云的形状逐渐变成一个与转轴相垂直的扁盘。这一形状对于解释太阳系行星大致处于同一平面是很重要的。

原恒星云经过一二十万年的进一步坍缩后，其中心区域因引力势能无法及时散去而造成的升温效应变得越来越显著，最终使得热运动产生的压强大到了能抗衡引力的程度。这时的原恒星云中心部分被称为原恒

星（protostar），它处于一种近似的流体静力平衡状态（hydrostatic equilibrium）。不过，这种近似的平衡状态虽能对原恒星本身的坍缩造成显著阻碍，却并不能阻止外侧物质继续向中心凝聚。这种凝聚过程被称为吸积过程（accretion），为这种过程“添砖加瓦”的外侧物质则被称为吸积盘（accretion disk，彩图15）。吸积过程对于恒星的成长是极其重要的，太阳刚刚成为原恒星时，它的质量仅为目前质量的1%左右，其余99%的物质全靠吸积过程来“侵吞”。

吸积过程从物理上讲是非常显而易见的，但由于被浓密的吸积盘所包围，起码在最初阶段是很难被观测到的，是星云说观测验证中为数不多的盲区之一。原恒星的吸积过程大约持续几十万年，当这一过程接近完成时，吸积盘上的物质已大都被原恒星所侵吞，使后者的质量大为壮大，并获得了一个新头衔：主序前星（pre-main sequence star）。

由于阻碍观测的吸积盘上的物质大都被侵吞，使得主序前星成为了一种能被直接观测到的天体。事实上，早在1852年，英国天文学家辛德（John Russell Hind，1823—1895）就在金牛座（Tauri）中发现了主序前星这种特殊阶段的天体，并将其编号为金牛座T星。当然，那时人们还并不知道此类天体在恒星演化过程中的地位，只是将其作为一颗亮度变化着的天体记录在册。如今，金牛座T星已成为了一大类天体的“代言人”，这类天体被统称为金牛座T型星（Tauri T star），它们是像太阳那样的小质量恒星的主序前星。^[5]

主序前星由于刚刚由星际云吸积而成，其“身材”在一定程度上还保持了星际云的“飘逸”，而远不如成熟恒星那样结实，它的表面温度比成熟恒星低得多，体积却极为庞大。除此之外，主序前星的另一个显著特点是有强劲的恒星风（对太阳来说就是太阳风）。事实上，原恒星的吸

积过程之所以终止，除了吸积盘上的物质基本用罄外，强劲的恒星风将残余物质吹离也是一个重要原因。有意思的是，主序前星所具有的表面温度低、体积庞大、恒星风强劲等特点都类似于后文将要介绍的红巨星（red giant star），只不过前者是恒星演化的早期阶段，后者则是晚期阶段，^[6]但这“一老一少”的遥相呼应，颇像人世间孩童与老人的相似之处。

主序前星的日常活动主要就是“瘦身”——在引力作用下进一步缓慢收缩，这种瘦身活动除了能优化“身材”外，一项最重要的功劳，就是为主序前星提供能量。我们在第8章中曾经提到过迈耶、亥姆霍兹、汤姆孙等人所主张的“引力说”。在那里，它是被当作失败假说来介绍的，但此刻我们却要为其恢复一点名誉，因为对于今天的太阳来说，它虽然是一个失败假说，但在主序前星那一“特定历史时期”它却是完全适用的。因为那时的太阳核心温度还没有高到能引发核聚变反应的程度，引力收缩确实是一种很重要的“替代能源”。

当主序前星的演化进行到后期时，随着核心温度的持续升高，一些“点火温度”较低的轻核将被相继点燃。比如当核心温度达到100万度时，最“易燃”的氘核将被点燃，核心温度达到300万、500万、600万度时锂核（ ^3Li ）、铍核（ ^4Be ）、硼核（ ^5B ）也将被相继点燃。这些轻核的聚变与引力收缩一起，为后期的主序前星提供能量。太阳作为主序前星的“瘦身”过程总共约持续几千万年至一亿年。最终，当核心温度达到800万度时，太阳一生的一个重要时刻来临了：它肚子里储备最丰富的氢元素终于被点燃了。这是太阳的“成人礼”，这时的太阳就正式步入了一生最稳定、并且持续时间很长的阶段——主序星（main sequence star）阶段。今天的太阳已经在这一阶段度过了46亿年的漫长时间，相对于预期寿命来说还只是“中年”而已。

主序星虽然是太阳一生最稳定的阶段，但在如此漫长的时间长河中，它的性质仍会发生缓慢变化。发生这种变化的主要原因，是氢核的含量经长期消耗后逐渐降低，导致核反应数量有所下降，从而使辐射压在与引力的抗衡中屈居下风。辐射压一旦屈居下风，太阳核心就会被引力所压缩。但这种压缩是暂时的，因为太阳核心一经压缩就会升温，而升温将使核聚变反应加速，重新产生出足够强的辐射压来抗衡引力。这种微妙的平衡机制在太阳内部持续起着作用，使太阳保持着总体的稳定。但这种核心温度的逐渐升高会传到外层，使太阳的表面温度也逐渐升高，并因此而逐渐膨胀。当然，这都是极缓慢的过程。研究表明，从主序星伊始到今天，太阳的表面温度大约升高了5%，半径大约膨胀了6%，光度则大约增加了40%。

很明显，这种升温 and 膨胀的趋势不仅存在于过去，也将持续存在于未来。计算表明，再过10亿年，太阳的光度将在今天的基础上再增加10%左右。这一变化将在地球上导致一系列连锁反应。首先是地表温度上升，接着是海洋蒸发加剧，然后是大气层中水蒸气的含量增加。再接下来则是最糟糕的一步，那就是温室效应。我们今天谈论温室效应时关注得最多的是二氧化碳，但实际上看似很无害的水蒸气才是最重要的温室效应气体（只不过它受人类活动的影响不如二氧化碳那样显著而已）。全球范围内水蒸气含量的增加将导致严重的温室效应。在比较悲观的估计中，整个海洋将因此而蒸发殆尽。当然，一定数量的地下水将得以留存，并在某些地方渗出地表，成为弥足珍贵的新水源。大气层中的水蒸气有时也会以暴雨的形式倾泻而下，形成昙花一现的湖泊。但地表的大部分地区将变成闷热的荒漠，多数生物将会绝种，幸存的物种将大都转入穴居生活。如果那时人类仍存在于地球上的话，或许将借助技术手段加入穴居的行列。

但那样的“好景”也是有期限的，随着太阳光度的继续增加，地球上的环境将继续恶化。在30亿年后的未来，地球的温室效应将彻底失控，那时的地球将会变得类似于今天的金星——那里的温度常年高达460°C，别说生物无法忍受，连熔点较低的金属都将被熔化。那时的地球有生物幸存的可能性将是微乎其微的。而如此凄惨的景象比起即将到来的更可怕的浩劫来说，仍然算不上什么。

在距今约40亿年之后，太阳一生的另一个重要时刻来临了——它核心区的氢燃料在经过了将近90亿年不间断的燃烧后终于到了“油尽灯枯”的一天。太阳从此告别了主序星阶段，步入老年恒星的行列。那时候，太阳的形态将再次发生重大变化，而离太阳最近两颗行星——水星和金星——则将迎来自己的末日。

由于氢燃料的耗尽，太阳核心区的下场只有一个，那就是坍缩。但这种坍缩导致的升温很快就会点燃核心区以外的一个壳层中的氢。因此，这时的太阳内部将发生两件事情：一件是氢燃料耗尽后的核心区在引力作用下发生坍缩；另一件则是核心区以外的一个壳层内的氢发生燃烧，成为新的核反应区。后者产生的高温高压将使整个太阳外层发生远比主序星阶段剧烈得多的膨胀，那时的太阳将变成一颗体积和光度都极为巨大，表面温度却因膨胀过于剧烈而下降到2 600 K左右的天体，那样的天体叫做红巨星^[7]。太阳在红巨星阶段将逗留几亿年。那么，变成红巨星后的太阳究竟有多大呢？一般认为，它的体积将达到目前体积的一千万倍，足以膨胀到目前的地球轨道之外。这一阶段的太阳虽然表面温度有所降低，但因表面积的巨大增加，总光度依然极为惊人，约为目前光度的两千倍以上。这是闪耀了近90亿年的太阳在谢幕前的最后演出。在这场演出中，离太阳最近的“群众演员”水星和金星将相继“殉职”，从太阳系行星行列中除名。

但我们脚下这颗地球的命运却比较微妙。

太阳的肚子虽然将膨胀到目前的地球轨道之外，但在这缓慢的膨胀过程中，比目前强劲得多的太阳风将带走大量的太阳物质（整个红巨星阶段被太阳风带走的质量有可能达到太阳总质量的 $\frac{1}{3}$ ）。太阳物质的损失将减弱太阳对地球的引力束缚，从而使地球公转轨道缓缓远离太阳。不过，这种远离是否足以使地球躲过烈火焚身的浩劫，学术界尚无定论。大家比较有共识的是：地球似乎恰好踩在自己的生死线上，靠近一点就是死，离远一点则是生。至于究竟是死还是生，不同的理论模型给出的结论不尽相同。

另外一点比较有共识的是，即便地球能侥幸躲过烈火焚身的浩劫，地球上的景象也将是不折不扣的地狱景象（彩图16）：殷红似血的太阳几乎占据整个天幕，两千多度的高温从近在咫尺的天空中疯狂地倾泻着热量，地表的很多物质将被熔化，岩浆般的洪流四处流淌，使这个曾经如此多姿多彩的生命乐园变得面目全非。但即便那样的浩劫也很可能只是将最终的毁灭稍稍延后一些而已，因为地球与比邻而居的太阳外层之间的潮汐作用将逐渐消耗地球轨道运动的能量，使地球的轨道逐渐向内缩减，重新投向死神的怀抱。

此时，太阳系外围天体的环境也将发生剧烈的变化，某些外行星卫星上的亘古寒冰将会融化，成为太阳系中新的海洋——当然，它们的成分与今天的地球海洋是截然不同的。

那么坍缩中的太阳核心又将如何呢？它会一直坍缩下去吗？答案是否定的。由于氢的耗尽，太阳核心的主要成分将变成氦，随着氢壳层的不断燃烧，越来越多新生成的氦将加入到坍缩中的氦核心上。而它的温度则随着坍缩的进行而不断升高。最终，当氦核心的质量达到太阳总质

量的45%以上时，它的坍缩将使温度升高到一亿度以上。这时候，很“耐热”的氦核终于也被点燃了，开始发生氦核聚变成碳核 ^{12}C ，以及氦核与碳核聚变成氧核 ^{16}O 的核聚变反应。这个点火过程几乎瞬息之间就能传遍整个氦核心，称为“氦闪”（Helium flash）。此后的太阳核心将进入最后一个燃烧阶段——氦燃烧阶段。

氦燃烧的一个很突出的特点，是它对温度的依赖性比氢燃烧还要敏感得多，哪怕核心温度只有2%的变化，也会导致光度的加倍或减半。由于这种敏感依赖性，太阳的光度和体积将会发生频繁的脉动，尚未被太阳风带走的太阳外层物质将会在这种脉动中惨遭抛弃。与氢燃烧相似，氦燃烧也将经历一个核心区的氦先耗尽，然后延烧至核心区以外的一个壳层上的过程，只不过整个历时比氢燃烧短得多，总共只有几千万年。

在太阳这整个晚年阶段中被太阳风及脉动抛射出去的太阳外层的物质，将在行将就木的太阳周围形成一片美丽的小星云，称为行星状星云（planetary nebula，彩图17）。这种星云早在18世纪时就被发现，因其在小型望远镜中形象与行星相类似而得名。但它们的真正身份则是在晚得多的时候才被“揭发”出来的。“揭发”的证据主要有两条：一是它们普遍在向外膨胀，说明它们是某种抛射过程的产物；二是在它们的中心发现了致密的核心，那样的核心与耗尽了核燃料后的恒星残骸完全相符。行星状星云中除了包含恒星外层那些未经燃烧的轻元素外，还包含一些被对流带到外层的碳、氧等恒星核聚变反应所生成的重元素。在遥远将来的某一天，它们也许将与宇宙中的其他物质汇集成新的大型星际云，并成为新一代恒星、行星，乃至生命的原材料。

对于太阳来说，那时的它已用尽了所有可资利用的核燃料，抛掉了

所有可以抛掉的外层物质，成为了一个富含氢燃烧产物——碳和氧——的光秃秃的内核。当然，其中还有大量的电子，因为太阳核心是电中性的，不可能只有原子核。这个内核的质量略多于目前太阳质量的一半，它初始时的温度极高，但却没高到足以点燃碳核与氧核的程度，燃烧了近百亿年的太阳核心洪炉就此永久性地熄了火。那么，此时的太阳用什么来抗衡引力呢？答案是：依靠一种新的压强电子简并压强（**electron degeneracy pressure**）。这种压强源于物理学上的一条著名原理，叫做“泡利不相容原理”，它表明像电子那样的粒子是一群极有“个性”的家伙，每个都想拥有一个独一无二的状态，这种“个性”导致的后果之一就是它们倾向于占据较大的空间。这就好比一群人如果个个都想拥有单独的房子，而不愿与人合租，他们所占据的空间就会较大。这种占据较大空间的倾向在宏观上的表现就是压强，这就是所谓的电子简并压强。

在太阳的核心永久熄火后，正是这种电子简并压强抗衡住了引力，使太阳处于一个稳定状态。这一状态极为致密，虽然拥有太阳的全部残留物质（约为地球质量的20万倍），体积却与地球相当。如果我们能从那时的太阳上舀下一汤匙物质的话，它将重达几吨！处于这一阶段的晚年恒星将在很长时间内维持很高的温度，发出炽热的白光，因此被称为白矮星（**white dwarf star**）。这种白矮星自20世纪初以来已被大量地发现，成为印证太阳演化理论最后一个环节的可靠证据。^[8]白矮星虽然炽热，但由于不再产生新的能量，最终将在冰冷的星际空间中逐渐冷却，由白变黑，成为一颗不再发光的冰冷天体——黑矮星，它的巨大引力场将是曾经生机勃勃的太阳系的最后墓碑。如果那时有什么智慧生物路过太阳的话，将只能从那巨大的引力场中发现它的存在，而它曾经抚育过的无数生灵，则早已成为过往烟云。

这就是太阳的归宿，也是我们太阳故事的尾声。^[9]当我们在本书的

开头遥望那光芒夺目的太阳时，哪怕有最高明的想象力，我们能想象出科学为我们勾勒出的如此宏大的图景吗？这种图景与最高明的想象相比也毫不逊色，但与单纯的想象截然不同，它的每一个结论都有直接或间接的证据，它的每一步推理都有严密可靠的逻辑，它用观测去印证，用逻辑来衔接，它所展现的是科学最动人心魄的力量和美感。

[1] 很难能可贵的是，在那个很多人视银河系为整个宇宙的年代，康德将星云视为了与银河系具有相似地位的所谓“岛宇宙”（island universe），与20世纪之后才确立起来的河外星系的观念不谋而合。

[2] 康德的《自然通史和天体论》有不只一种中译本，较新的译本收录于《康德著作全集》第一卷（李秋零主编，中国人民大学出版社2003年版）。不过有必要提醒有兴趣参阅该译本的读者注意的是：该书第334页将康德的这一臆想错译成了含义截然相反的“他们的居处离太阳越近，他们就越优秀、越完善”。

[3] 与之相对应地，HI区和HII区星际云通常被合称为漫射云（diffuse cloud）。不过“漫射云”这个译名（参阅中国天文学会天文学名词审定委员会的“天文学名词”网站）我个人认为不是很妥帖，因为“漫射”一词给人一种较明显的动态感觉，与此处的实际含义有一定出入，不如译成“弥散云”更为合适。另外要指出的是，此处所谓的“稠密”和“弥散”是相对的，与我们日常所接触的气体相比，哪怕稠密云也是“弥散”得一塌糊涂的东西，密度不到海平面空气的 10^{16} 分之一，甚至比一般实验室里的高真空还要“弥散”得多。

[4] 比如HI区星际云的金斯质量往往高达100 000个太阳质量以上，数量要比分子云稀少得多。不过稀少归稀少，那样的星际云毕竟还是有

的，因此确实有一部分恒星的形成被认为是始于HI区星际云的坍缩。

[5]除金牛座T型星外，主序前星还有另外两个主要类别：猎户座FU型星（FU Orionis star）和赫比格Ae/Be型星（Herbig Ae/Be star），其中前者一般被认为是金牛座T型星的早期演化阶段，而后者则是大质量恒星的主序前星。

[6]这一点也正是对主序前星与红巨星进行观测区分的关键所在，因为恒星演化的早期阶段与经历了核聚变反应后的晚期阶段在元素丰度上存在很大差异，这种差异可以用光谱学手段予以分辨。

[7]红巨星作为恒星晚年演化阶段的想法是20世纪30年代之后确立起来的。但有意思的是，英国小说家威尔斯（H. G. Wells, 1866—1946）在1895年发表的著名科幻小说《时间机器》（The Time Machine）中有一段关于太阳未来的描述非常类似于红巨星：“我被地球的命运之谜所吸引，停停走走，一步千年地迈进着，怀着奇异的兴趣注视着西边天幕上的太阳变得越来越大，越来越暗淡……最终，在三千多万年之后，太阳的巨大红色圆面遮蔽了将近十分之一的暗淡天空。”不过，撇开字面上的相似性不论，威尔斯对红巨星的真正威力显然没什么认识，在他的小说中，那时的地球上仍存在着地衣、苔藓及很多穴居的生物。

[8]白矮星只是像太阳这样的小质量恒星的归宿，而非恒星的唯一归宿。质量远大于太阳的恒星将会坍缩成更致密的中子星，甚至成为黑洞。相应的，它们的核聚变反应也不会终止于氦燃烧，而将经历若干重核的燃烧，并且它们在“临死之前”还将发生威力惊人的大爆炸——超新星爆发（supernova）。

[9]当然，这还不是太阳最终、最遥远的命运，后者是与整个宇宙的命运息息相关的，所涉及的因素包括宇宙学常数的大小、宇宙的开放与封闭等。那些话题由于远远超出了本书的范围，就不在此处介绍了。

附录 太阳档案

平均直径(mean diameter)	139.2 万千米
椭率(ellipticity)	0.000 05
质量(mass)	$1.989\ 1\times 10^{27}$ 吨
平均密度(mean density)	1 408 千克/立方米
表面重力加速度(surface gravity)	274.0 米/秒 ²
表面逃逸速度(escape velocity)	617.6 千米/秒
目视星等(visual magnitude)	-26.74
绝对星等(absolute magnitude)	4.83
光谱类型(spectral type)	G2 V
光度(luminosity)	3.846×10^{26} 瓦特
表面有效温度(effective temperature)	5 778 度(绝对温标)
中心温度(central temperature)	1 571 万度(绝对温标)
中心密度(central density)	162 200 千克/立方米
中心压强(central pressure)	2 477 亿帕斯卡
地球平均距离(mean distance from Earth)	1.496 亿千米
银心平均距离(mean distance from milky way core)	26 000 光年
绕银心公转周期(galactic period)	2.25 亿~2.50 亿年
自转周期(rotation period)	16°: 25.38 日 赤道: 25.05 日 两极: 34.3 日
自转倾角(obliquity)	与黄道面: 7.25° 与银道面: 67.23°

人名索引

A

阿波罗尼斯 (Apollonius)

阿尔文 (Hannes Alfvén)

阿基米德 (Archimedes)

阿奎纳斯 (Thomas Aquinas)

阿里斯塔克斯 (Aristarchus)

阿斯顿 (Francis William Aston)

阿西莫夫 (Isaac Asimov)

埃拉托斯特尼 (Eratosthenes)

埃文斯 (John Evans)

艾伦菲斯特 (Paul Ehrenfest)

爱丁顿 (Arthur Eddington)

爱因斯坦 (Albert Einstein)

奥本海默 (Robert Oppenheimer)

奥普泽（Theodor von Oppolzer）

奥希亚德（Andreas Osiander）

B

巴克尔（John Bahcall）

柏拉图（Plato）

贝克勒尔（Henri Becquerel）

贝纳（Henri Bénard）

贝塞尔（Friedrich Bessel）

贝特（Hans Bethe）

本生（Robert Bunsen）

比尔曼（Ludwig Biermann）

玻尔（Niels Bohr）

玻尔兹曼（Ludwig Boltzmann）

玻意耳（Robert Boyle）

布鲁诺（Giordano Bruno）

C

查德维克 (James Chadwick)

查普曼 (Sydney Chapman)

D

戴森 (Frank Dyson)

戴维斯 (Raymond Davis Jr.)

德伯纳 (Franz-Ludwig Deubner)

德拉鲁 (Warren De la Rue)

德雷尔 (Sidney Drell)

德西特 (Willem de Sitter)

狄拉克 (Paul Dirac)

笛卡儿 (René Descartes)

第谷 (Tycho Brahe)

杜瓦尔 (Tom Duvall)

F

费恩曼 (Richard Feynman)

费米 (Enrico Fermi)

夫浪禾费（Joseph von Fraunhofer）

弗雷泽（Edward Frazier）

弗洛因德利希（Erwin Freundlich）

福塞特（Eric Fossat）

G

伽利略（Galileo Galilei）

伽莫夫（George Gamow）

哥白尼（Nicolaus Copernicus）

格雷克（Gerard Grec）

格里波夫（Vladimir Gribov）

格罗斯曼（Marcel Grossmann）

H

哈维（Geoffrey Harvey）

哈维（Jaqck Harvey）

海森伯（Werner Heisenberg）

亥姆霍兹（Hermann von Helmholtz）

赫茨普龙 (Ejnar Hertzsprung)

赫歇耳 (John Herschel)

赫歇耳 (William Herschel)

黑尔 (George Hale)

黑尔斯 (Thomas Hales)

亨德森 (Thomas Henderson)

胡克 (Robert Hooke)

怀特黑德 (Alfred Whitehead)

霍尔顿 (Gerald Holton)

J

基尔霍夫 (Gustav Kirchhoff)

教宗保罗三世 (Pope Paul III)

教宗厄本八世 (Pope Urban VIII)

教宗约翰·保罗二世 (Pope John Paul II)

金斯 (James Jeans)

K

卡文迪许 (Henry Cavendish)

开普勒 (Johannes Kepler)

坎贝尔 (William Campbell)

康德 (Immanuel Kant)

考恩 (Clyde Cowan)

克莱因 (Felix Klein)

克里奇菲尔德 (Charles Critchfield)

孔德 (Auguste Comte)

库兹明 (Vadim Kuzmin)

L

拉普拉斯 (Pierre-Simon Laplace)

莱顿 (Robert Leighton)

莱因斯 (Frederick Reines)

朗道 (Lev Landau)

朗诺 (Lon Nol)

里根 (Ronald Reagan)

卢瑟福 (Ernest Rutherford)

罗素 (Bertrand Russell)

罗素 (Henry Russell)

罗兹 (Edward Rhodes)

洛克耶 (Norman Lockyer)

洛伦兹 (Hendrik Lorentz)

M

马克西米利安一世 (Maximilian I)

迈耶 (Julius von Mayer)

米克耶夫 (Stanislav Mikheyev)

莫塞莱 (Henry Moseley)

默里 (Andrew Murray)

N

能斯特 (Walther Nernst)

牛顿 (Isaac Newton)

O

欧多克斯 (Eudoxus)

P

庞蒂科夫 (Bruno Pontecorvo)

泡利 (Wolfgang Pauli)

佩恩 (Cecilia Payne)

佩林 (Jean Perrin)

普朗克 (Max Planck)

普朗特 (Ludwig Prandtl)

普耶特 (Claude Pouillet)

Q

钱德拉塞卡 (Subrahmanyan Chandrasekhar)

S

塞曼 (Pieter Zeeman)

塞奇 (Angelo Secchi)

施瓦布 (Heinrich Schwabe)

史瓦西 (Karl Schwarzschild)

斯顿夫 (Carl Stumpf)

斯米诺夫 (Alexei Smirnov)

斯塔克 (Johannes Stark)

斯特藩 (Joseph Stefan)

索德纳 (Georg von Soldner)

索迪 (Frederick Soddy)

T

泰勒 (Edward Teller)

汤川秀树 (Hideki Yukawa)

汤姆孙 (J. J. Thomson)

汤姆孙 (William Thomson)

托勒密 (Ptolemy)

W

瓦耳勒 (Jules Violle)

外尔 (Hermann Weyl)

威尔斯 (H. G. Wells)

维恩 (Wilhelm Wien)

魏茨泽克 (Carl von Weizsäcker)

沃芬斯坦 (Lincoln Wolfenstein)

沃拉斯顿 (William Wollaston)

沃特斯顿 (John Waterston)

乌尔里克 (Roger Ulrich)

X

西尔维斯坦 (Ludwik Silberstein)

希尔伯特 (David Hilbert)

希罗多德 (Herodotus)

希帕克 (Hipparchus)

小柴昌俊 (Masatoshi Koshihara)

辛德 (John Russell Hind)

Y

亚里士多德 (Aristotle)

英费尔德（Leopold Infeld）

Z

詹森（Pierre Janssen）

术语索引

GALLEX

g模

Homestake

MSW效应

p模

SAGE

SNO

γ 射线

《费恩曼物理学讲义》

《关于两大世界体系的对话》

《科学与当代世界》

《圣经》

《时间机器》

《世界的和谐》

《天体运行论》

《物理学的进化》

《宇宙体系论》

《自然通史和天体论》

《自然哲学的数学原理》

A

阿尔文波

奥尔特云

B

白矮星

半影

北极星

贝利珠

贝纳胞

本轮

本生灯

本影

本征态

标准模型

标准太阳模型

表观视运动

表观太阳日

C

差旋层

超级神冈探测器

超新星

传导

磁通量管

D

大麦哲伦星云

等离子体

等效原理

地磁暴

地心说

电荷守恒定律

电子

电子简并压强

定天镜

对流

对流区

对流坍塌

多普勒效应

惰性中微子

F

发电机机制

分子云

丰度

夫琅禾费线

辐射

辐射区

G

格林尼治天文台

光度

光谱

光谱分析

光谱仪

光球层

广义相对论

过渡区

H

氦闪

核聚变

赫比格Ae/Be型星

赫罗图

黑矮星

黑体辐射

黑子

恒星日

恒星月

红巨星

混合程理论

混合矩阵

J

基尔霍夫电路定律

基尔霍夫热辐射定律

极光

降交点

交点的退行

交点年

交点线

截止频率

介子

金牛座T星

金斯不稳定性

金斯质量

进动

近点月

经院哲学家

均轮

K

开普勒猜想

科氏力

可栖息带

L

拉格朗日点

莱顿天文台

里克天文台

量子理论

猎户座FU型星

M

蒙德极小期

米粒

冕洞

N

能量守恒定律

宁静太阳

牛顿冷却定律

P

泡利不相容原理

偏心等距点

平均太阳日

Q

切连科夫辐射

全环食

全球振荡监测网

R

日环食

日冕

日冕物质抛射

日偏食

日球层

日球层顶

日全食

日食

日心说

日震学

S

塞曼效应

三角视差法

三棱镜

色球层

色散

沙罗序列

沙罗周期

嬗变

神冈探测器

升交点

食季

食年

史瓦西解

史瓦西判据

束后本影

朔望月

斯波勒定律

斯塔克效应

斯特藩-玻尔兹曼常数

斯特藩-玻尔兹曼定律

T

太阳风

快太阳风

慢太阳风

太阳风暴

太阳和日球层探测器

太阳活动

太阳活动极大年

太阳活动极小年

太阳日

太阳中微子单位

太阳中微子问题

太阳周期

碳氮氧循环

天体照相机

湍流

W

万有引力常数

维恩位移定律

温度最低层

温室效应

五分钟振荡

X

吸积过程

吸积盘

狭义相对论

谐振腔

星际云

星云假说

星云说

行星凌日

行星状星云

Y

阳历

耀斑

纤耀斑

阴历

原恒星

原恒星云

月相

Z

针状物

质能关系式

质子

质子-质子链

中微子

太阳中微子

电子中微子

μ 子中微子

τ 子中微子

中微子振荡

中性流反应

中子

终端激波

周年视运动

周日视运动

主序前星

主序星

参考文献

[1] Armitage A. Copernicus and Modern Astronomy [M] . New York: Dover Publications Inc, 2004.

[2] Bartusiak M. Archives of the Universe: 100 Discoveries That Transformed Our Understanding of the Cosmos [M] . New York: Vintage Books, 2004.

[3] Bhatnagar A, Livingston W. Fundamentals of Solar Astronomy [M] . Berlin: Springer-Verlag, 2010.

[4] Bilenky S. Introduction to the Physics of Massive and Mixed Neutrinos [M] . Singapore: World Scientific Publishing Co Pte Ltd, 2005.

[5] Celnikier L M. Find a Hotter Place! A History of Nuclear Astrophysics [M] . Singapore: World Scientific Publishing Co Pte Ltd, 2006.

[6] Chandrasekhar S. 爱丁顿——当代天体物理学家 [M] . 吴智仁, 等, 译. 上海: 远东出版社, 1992.

[7] Chaplin W J. Music of the Sun: The Story of Helioseismology [M] . London: Oneworld Publications, 2006.

[8] Clerke A M. A Popular History of Astronomy during the

Nineteenth Century [M] . London: Adam & Charles Black, 1893.

[9] Crelinsten J. Einstein's Jury: The Race to Test Relativity [M] . Princeton: Princeton University Press, 2006.

[10] Dwivedi B N. Dynamic Sun [M] . Cambridge: Cambridge University Press, 2003.

[11] Enghag P. Encyclopedia of the Elements [M] . Hoboken: Wiley-VCH Verlag GmbH & Co KGaA, 2004.

[12] Foukal P V. Solar Astrophysics [M] . Hoboken: Wiley-VCH Verlag GmbH & Co KGaA, 2004.

[13] Franklin A. Are There Really Neutrinos [M] . New York: Perseus Books, 2001.

[14] Fukugita M. Yanagida T. Physics of Neutrinos and Applications to Astrophysics [M] . Berlin: Springer-Verlag Berlin Heidelberg, 2003.

[15] Galileo G (author) . Finocchiaro M (ed) . The Essential Galileo [M] . Indianapolis: Hackett Publishing Company, Inc, 2004.

[16] Giunti C. Kim C W. Fundamentals of Neutrino Physics and Astrophysics [M] . Oxford: Oxford University Press, 2007.

[17] Golub L. Pasachoff J M. Nearest Star: The Surprising Science of Our Sun [M] . Cambridge: Harvard University Press, 2002.

[18] Harrington P S. Eclipse: The What, Where, When, Why,

and How Guide to Watching Solar and Lunar Eclipses [M] . New York: John Wiley & Sons, Inc, 2008.

[19] Holton G. 物理科学的概念和理论导论 [M] . 张大卫, 等, 译. 北京: 高等教育出版社, 1983.

[20] Hoskin M. The Cambridge Illustrated History of Astronomy [M] . Cambridge: Cambridge University Press, 1997.

[21] Isaacson W. Einstein: His Life and Universe [M] . New York: Simon & Schuster, 2008.

[22] Jackson M W. Spectrum of Belief: Joseph von Fraunhofer and the Craft of Precision Optics [M] . Cambridge: The MIT Press, 2000.

[23] Kennefick D. Not Only Because of Theory: Dyson, Eddington and the Competing Myths of the 1919 Eclipse Expedition [J/OL] . <http://arXiv.org/abs/0709.0685v2>.

[24] Kline M. 古今数学思想 (第一册) [M] . 张理京, 张锦炎, 江泽涵, 译. 上海: 上海科学技术出版社, 2002.

[25] Lang K R. The Cambridge Encyclopedia of the Sun [M] . Cambridge: Cambridge University Press, 2009.

[26] Lang K R. The Sun from Space [M] . Berlin: Springer-Verlag Berlin Heidelberg, 2009.

[27] Littmann M. Espenak F, Willcox K. Totality: Eclipses of the Sun [M] . Oxford: Oxford University Press, 2001.

- [28] Lowe A. Neutrino Physics & The Solar Neutrino Problem [J/OL] . <http://arXiv.org/abs/0907.3658v1>.
- [29] Mobberley M. Total Solar Eclipses [M] . Berlin: Springer Science + Business Media, LLC, 2007.
- [30] Motz L. The Story of Astronomy [M] . New York: Perseus Publishing, 1995.
- [31] Pais A. Subtle is the Lord: The Science and the Life of Albert Einstein [M] . Oxford: Oxford University Press, 2005.
- [32] Rieutord M. The Sun's Supergranulation [J] . Living Rev Solar Phys, 2010, 7.
- [33] Russell B. Religion and Science [M] . Oxford: Oxford University Press, 1997.
- [34] Steel D. Eclipse [M] . The Joseph Henry Press, 2001.
- [35] Stix M. The Sun: an Introduction [M] . Berlin: Springer-Verlag Berlin Heidelberg, 2004.
- [36] Tassoul J. Tassoul M. A Concise History of Solar and Stellar Physics [M] . Berlin: Princeton University Press, 2004.
- [37] Taylor S R. Solar System Evolution: A New Perspective [M] . Cambridge: Cambridge University Press, 2001.
- [38] Thomas I (trans) . Selections Illustrating the History of Greek

Mathematics II [M] . London: William Heinemann Ltd, 1941.

[39] Ward P, Brownlee D. Rare Earth: Why Complex Life Is Uncommon in the Universe [M] . Berlin: Springer, 2003.

[40] Whitehouse D. The Sun: A Biography [M] . New York: John Wiley & Sons, Ltd, 2004.

[41] Zirker J B. Journey from the Center of the Sun [M] . Princeton: Princeton University Press, 2004.

[42] Zirker J B. The Magnetic Universe [M] . Baltimore: The Johns Hopkins University Press, 2009.

[43] 黄润乾. 恒星物理 [M] . 北京: 科学出版社, 2000.

[44] 林元章. 太阳物理导论 [M] . 北京: 中国科学技术出版社, 2006.

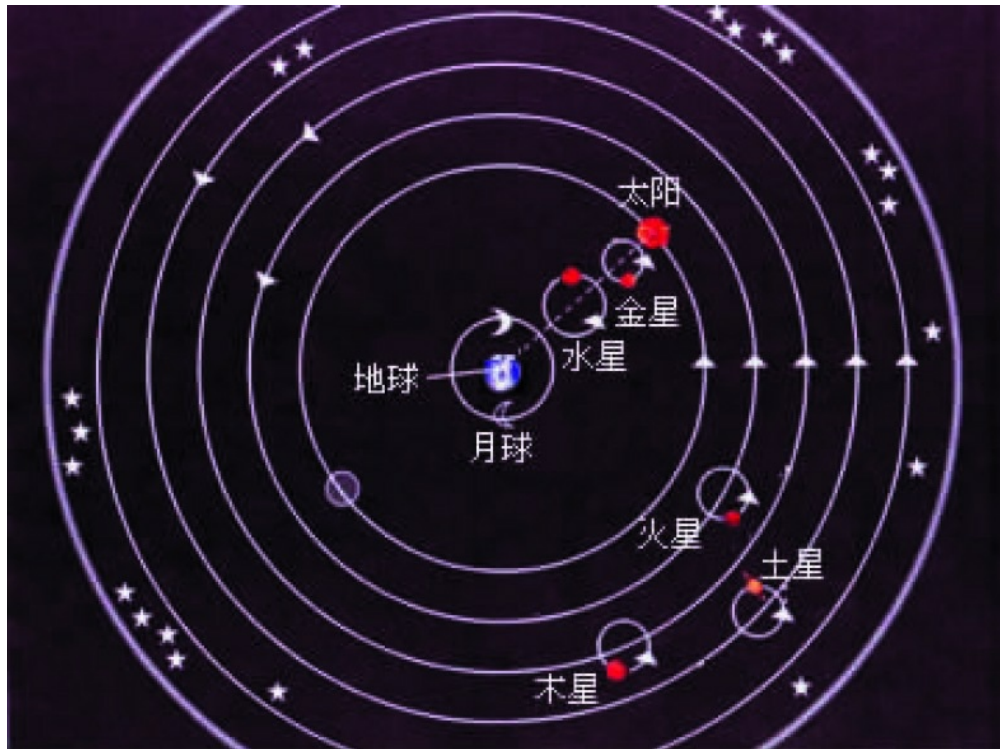
[45] 张昌明, 肖耐园. 天文学教程 (上册) [M] . 北京: 高等教育出版社, 1987.



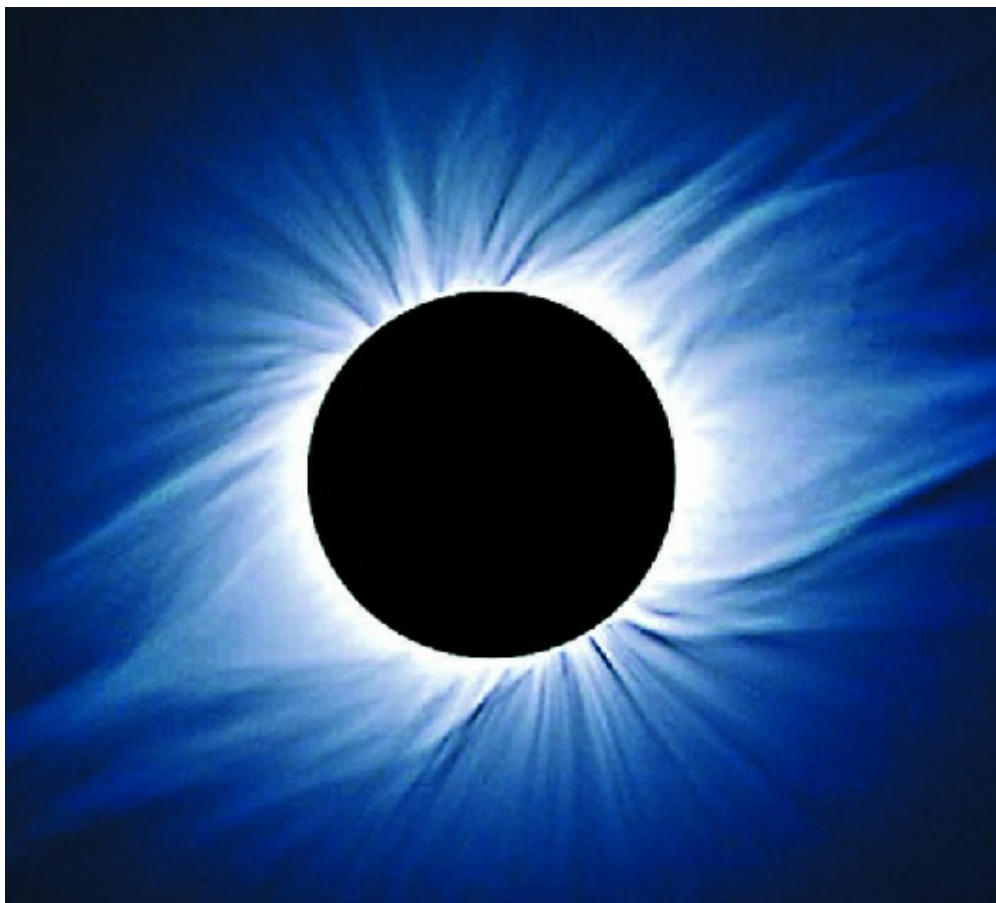
彩图1 夕阳下的古希腊神庙遗址



彩图2 月相的变化



彩图3 简化版的地心说模型



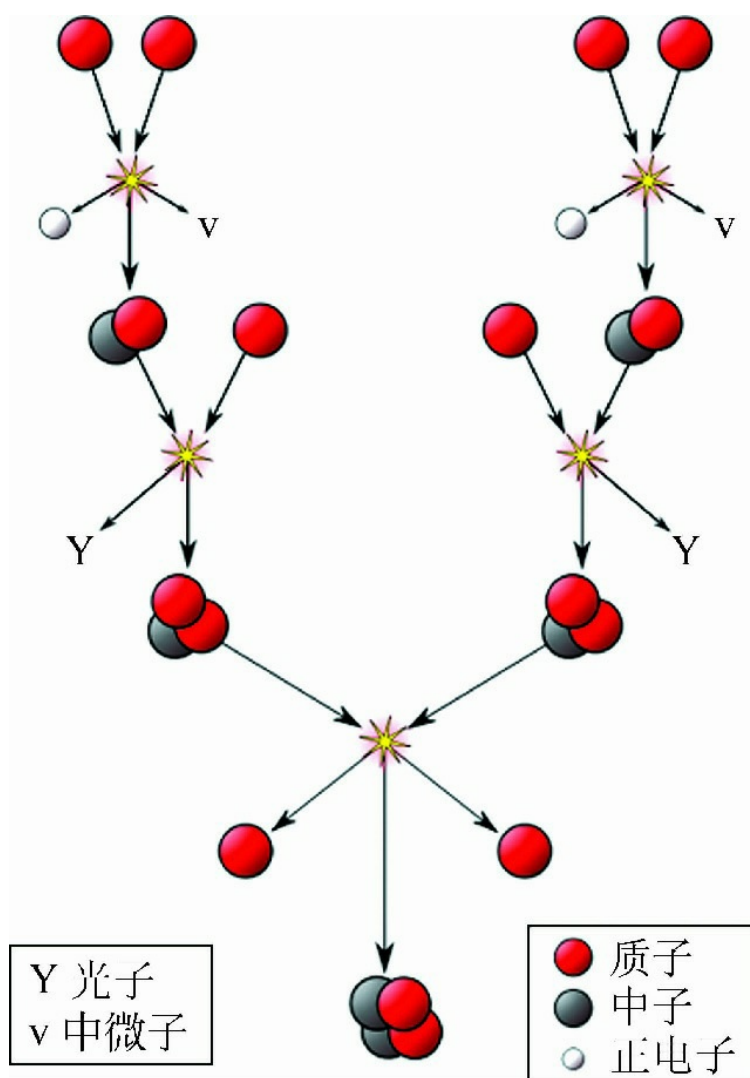
彩图4 日全食



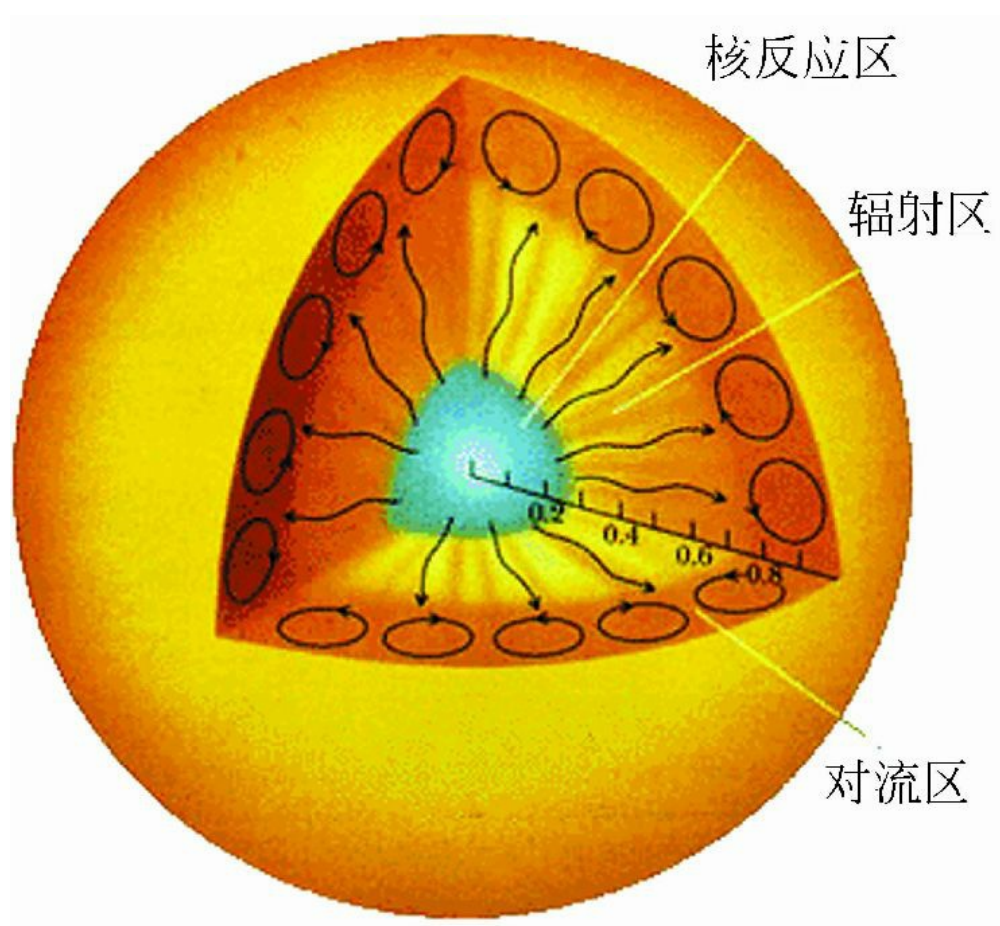
彩图5 光的色散



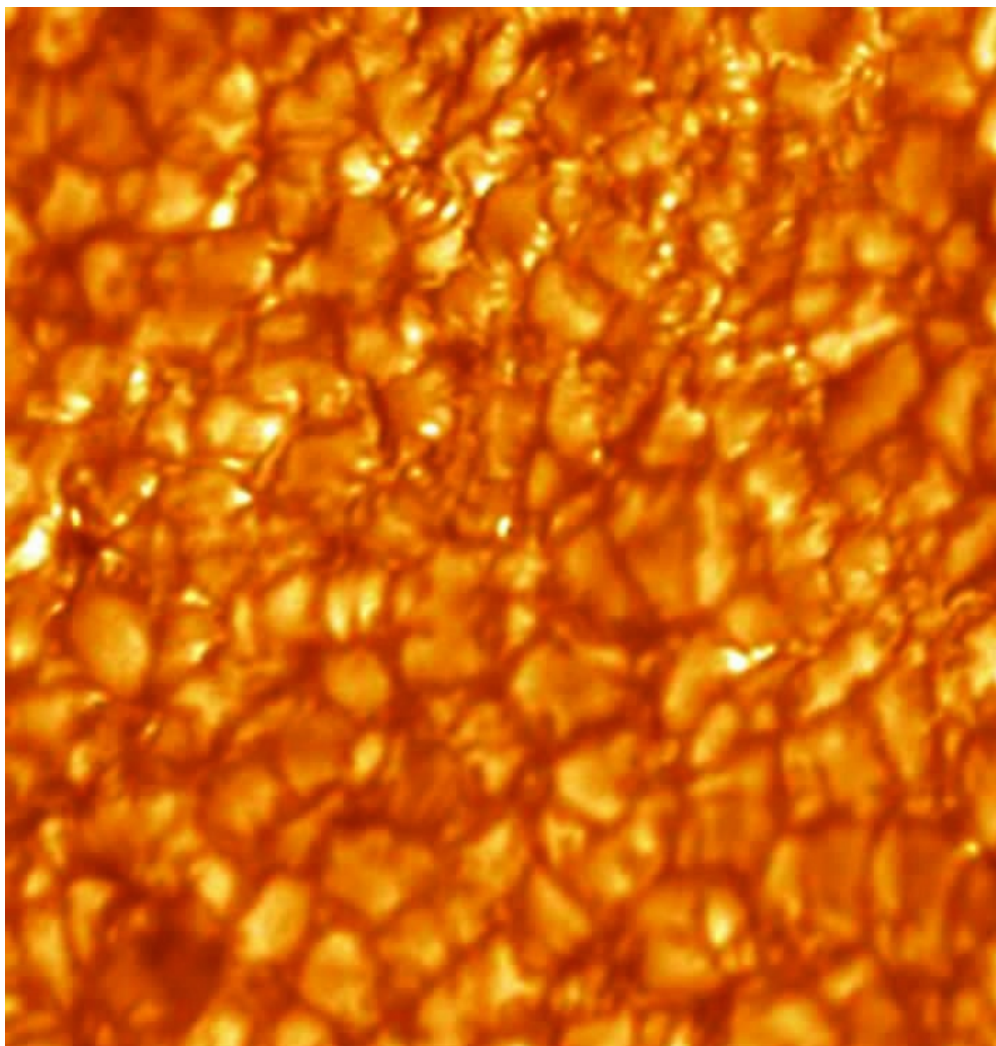
彩图6 纪念夫琅禾费诞辰两百周年的邮票



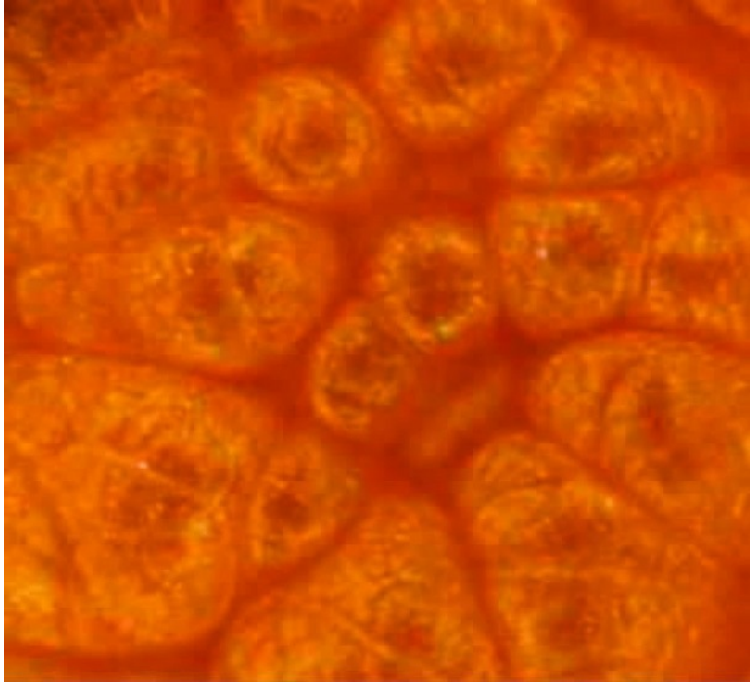
彩图7 第一类质子-质子链



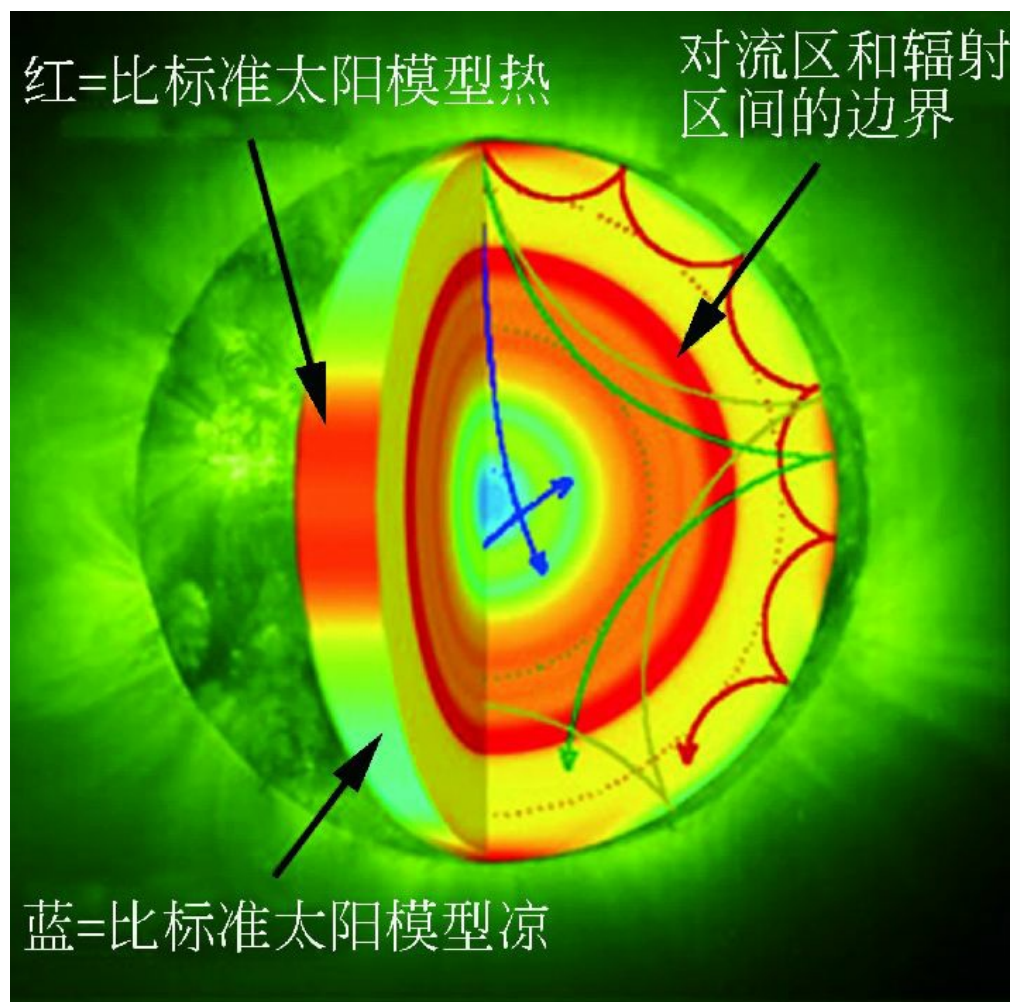
彩图8 太阳的内部结构



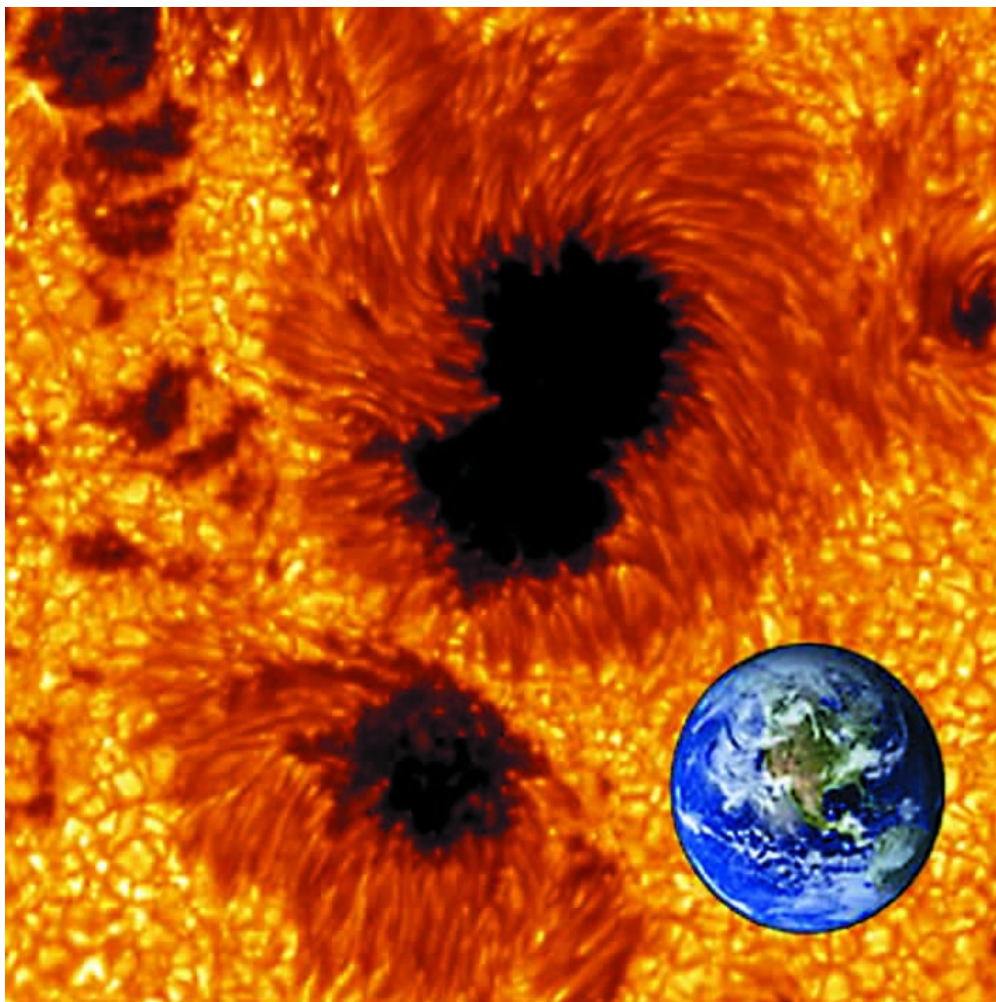
彩图9 太阳表面的“米粒”



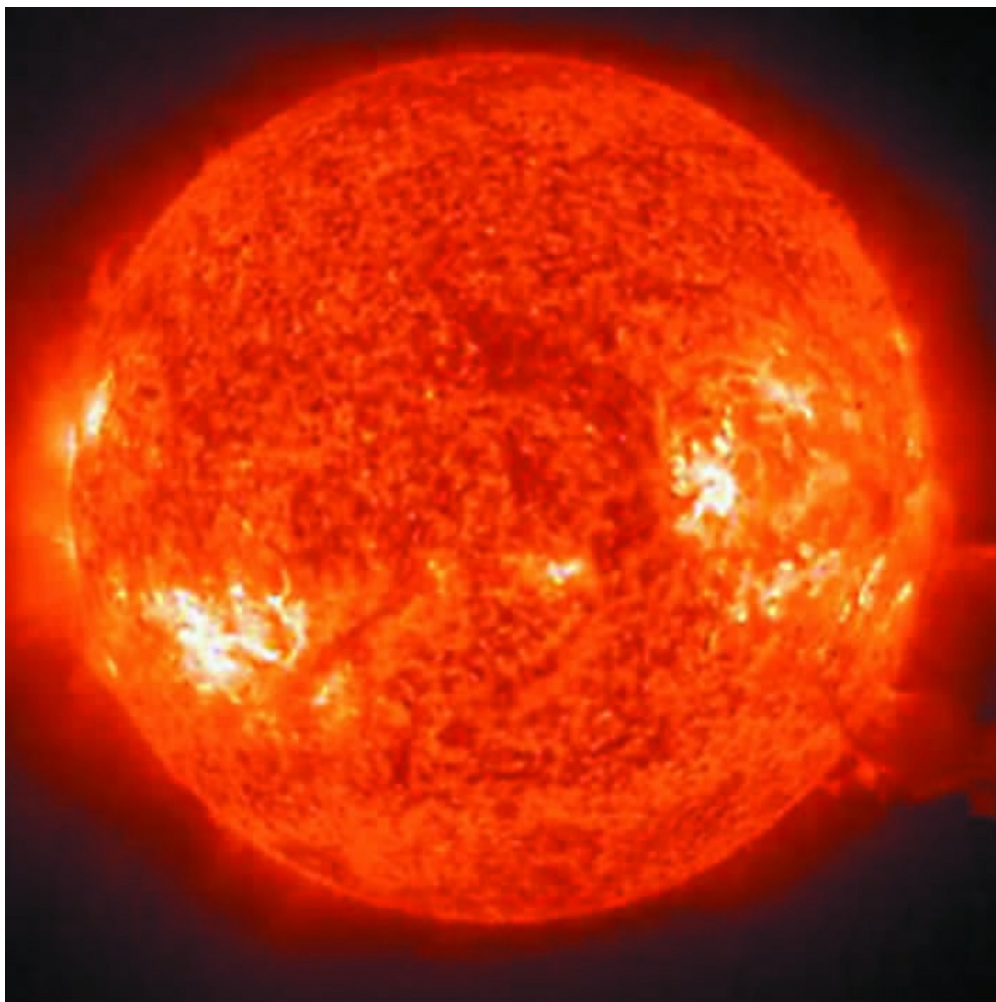
彩图10 普通流体中贝纳胞



彩图11 太阳内部的声波模式



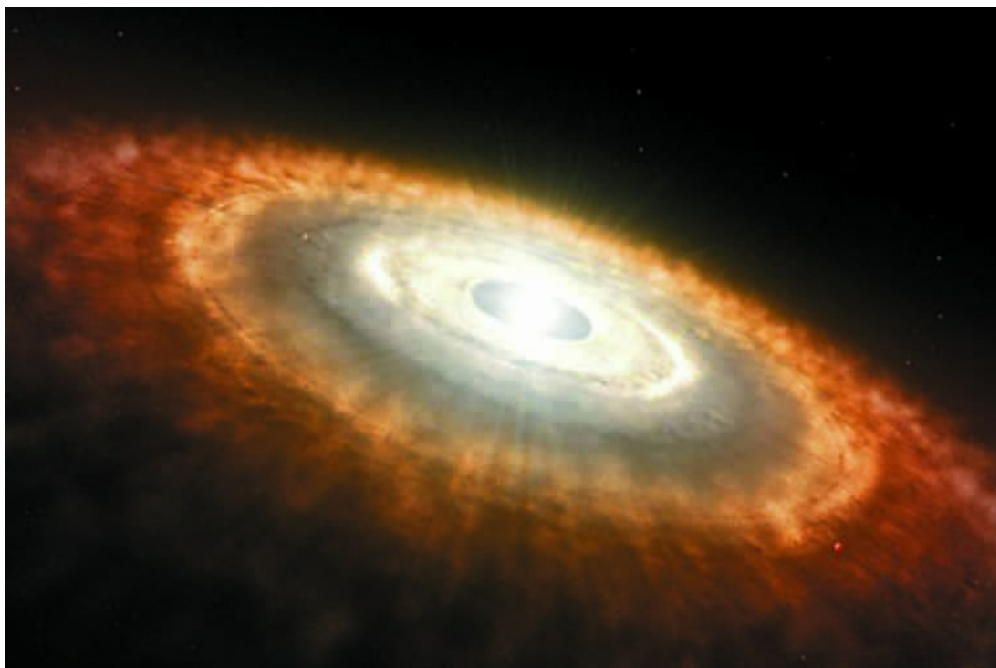
彩图12 太阳黑子（右下角的对照物为地球）



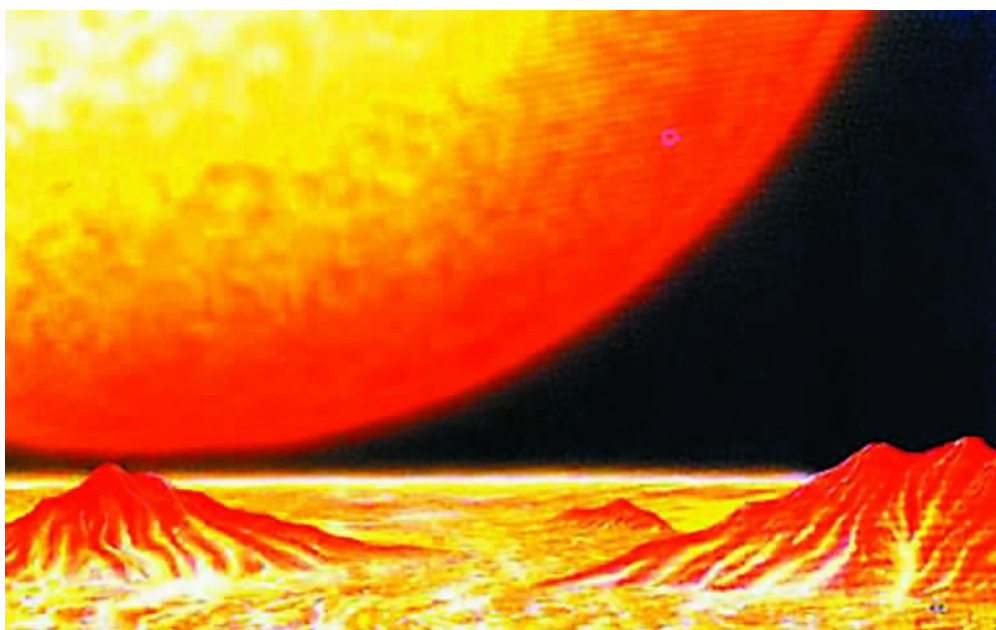
彩图13 太阳耀斑



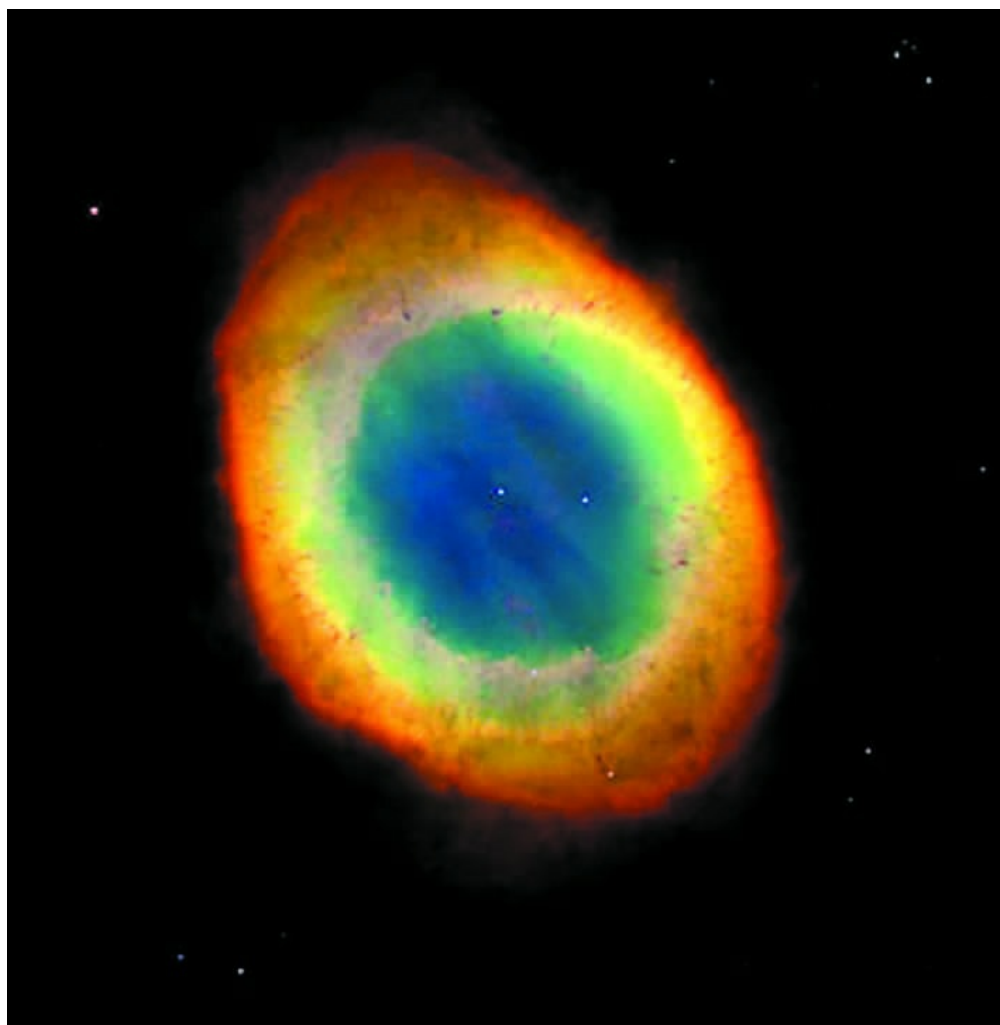
彩图14 离我们较近的分子云：猎户星云 (Orion Nebula)



彩图15 原恒星与吸积盘想象图



彩图16 太阳变成红巨星后的地球想象图



彩图17 行星状星云

理解科学丛书

SUNRISE

A Story of
the Quantum Theory and Relativity

日出

量子力学与相对论

张轩中 黄宇傲天◎著

在20世纪初期乌云密布的物理学天空中，
量子力学和相对论如何喷薄而出？

讲述一个震撼人心的英雄时代。
在物理学精英的传奇故事中，
挖掘产生新思想的源动力。

清华大学出版社

目 录

[序1](#)

[序2](#)

[序3](#)

[第一部分 量子力学](#)

[1 孤儿：两个钻石王老五](#)

[2 监狱里的群论](#)

[3 爱因斯坦与玻耳兹曼，马赫](#)

[4 写清楚熵公式的人](#)

[5 富二代德布罗意：凌晨旧戏](#)

[6 薛定谔：遗情书](#)

[7 十年前的玻尔](#)

[8 物理思想集大成者：爱因斯坦](#)

[9 犀利哥](#)

[10 海森堡：日出](#)

[11 交换游戏](#)

[12 物理学生市：矩阵的运动方程](#)

[13 科莫湖畔的费米](#)

[14 第五届索尔维会议](#)

[15 狄拉克矩阵：相对论与量子力学的婚姻](#)

[16 海鸥，马约拉纳](#)

[17 朗道：苏联之子](#)

[18 广岛之吻](#)

[19 原子弹研发与物理学熊市](#)

[20 枪手，猫论](#)

[21 在战后的废墟上](#)

[第一部分附录](#)

[第二部分 相对论](#)

[22 乡下的月光](#)

[23 牛顿引力的高级版本：拉普拉斯方程](#)

[24 三体问题](#)

[25 惯性参考系](#)

[26 伯尔尼克拉姆大街49号](#)

[27 公务员的奋斗](#)

[28 三年半的沉默](#)

[29 朗之万：双胞胎悖论](#)

[30 广义相对论](#)

[31 美国空军的研究员](#)

[32 贝肯斯坦和霍金](#)

[33 宇宙学](#)

[第二部分附录](#)

[参考文献](#)

[跋](#)


[返回总目录](#)

作者简介

张轩中，北京师范大学物理系理论物理专业毕业，从事离子阱与质谱仪器研究。著有《相对论通俗演义》，曾在《北京青年报》等媒体发表多篇文艺与科学评论。

SUNRISE

A Story of
the Quantum Theory and Relativity

 理解科学丛书

SUNRISE
A Story of
the Quantum Theory and Relativity

日出

量子力学与相对论

张轩中 黄宇傲天◎著

清华大学出版社
北京

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

日出：量子力学与相对论 / 张轩中，黄宇傲天著. —北京：清华大学出版社，2013

（理解科学丛书）

ISBN 978-7-302-33947-2

I. ①日... II. ①张...②黄... III. ①量子力学—青年读物②量子力学—少年读物③相对论—青年读物④相对论—少年读物 IV. ①0413.1-49②0412.1-49

中国版本图书馆CIP数据核字（2013）第223507号

责任编辑：朱红莲

封面设计：蔡小波

责任校对：刘玉霞

责任印制：刘海龙

出版发行：清华大学出版社

网址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地址：北京清华大学学研大厦A座 邮 编：100084

社总机：010-62770175 邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市金元印装有限公司

经 销：全国新华书店

开 本：165mm×240mm 印 张：20.5 字 数：291千字

版 次：2013年12月第1版 印 次：2013年12月第1次印刷

产品编号：054846-01



海森堡在海边看日出 绘画：杨千



作者之一张轩中和霍金在一起

谨以此书
献给青年物理学家张宏宝博士

序1

本书是张轩中学子继《相对论通俗演义》之后的又一力作。轩中从中学时代起就是一位对文学和科学都充满兴趣，而且有志于创造的青年。他先受张爱玲、徐志摩等人作品的影响而关注文学，后又受康德、爱因斯坦、杨振宁等人的影响转而关注科学，最终选择了物理专业，进入北京师范大学物理系学习，加入该校的相对论小组深造。这个小组是全国最有影响力的相对论研究团队之一。在刘辽教授和梁灿彬教授的领导下，为年轻人打下较为深厚的物理基础和数学基础，并引导他们走向引力理论、时空理论、宇宙学和黑洞物理的研究前沿。

轩中深受梁灿彬教授的影响，对理论物理和微分几何充满兴趣，并在马永革教授的指导下完成了硕士论文。在本科和研究生阶段，轩中博览群书，把自己的视野从相对论扩展到量子论及物理学史等领域。

轩中最值得注意的优点是始终想做事情，不想虚度此生。在各种尝试之后，他终于摸索到了适合自己的创造领域，即选择科学和文学的结合部，把掌握的物理学前沿知识与自己的写作能力结合起来，以既科学又文学的方式把爱因斯坦、霍金等近代学者研究的前沿知识及成功者的创新之路介绍给广大青年。

我怀着欣喜的心情读完了本书的电子稿，大有后生可畏又可爱的感觉。这本书和几年前发表的《相对论通俗演义》一样，既保证了科学内容的正确，历史梗概的真实，又带有适当的文学色彩，而且使用的是年轻人熟悉喜爱的语言，可读性很强。《相对论通俗演义》一书不但受到了青年读者的喜爱，也受到了许多理论物理专家的好评，相信这本书同样会得到较高的评价。

我们中华民族是一个伟大的民族，从春秋战国到明朝中叶，站在人类文明的前列差不多2000年，只是在近代落后了，而且挨打了。经过全

国人民的浴血奋斗，今天的中国终于又站起来了，经济得到了飞速的发展，中华民族再次奔向人类文明的最前列。在目前经济危机的阴影笼罩全球的情况下，中国是最有活力、最充满希望的国家，全世界人民都在注意正在重新崛起的中华民族，对于当代中国青年来说，真是任重道远。

邓小平同志说，“科学技术是第一生产力”。中国经济的可持续发展依赖于科技创新和教育的普及提高。几十年来，中国的教育取得了长足进展，不仅基本普及了中、小学教育，而且大学生能占到同龄人的三分之一，这是震撼世界的成就，是中国可持续发展的强大动力之一。

我们的教育质量，从知识水平和计算能力来说是不低的，但也有严重缺陷，这就是中国学生的创新动力和创新能力都不足。

应该使年轻人明白，500年的自然科学史表明，青年是科学发现的主力军。伽利略25岁被誉为“当代的阿基米德”。牛顿23~25岁期间完成了他一生最重要的科学发现。莱布尼茨27岁发明微积分，伽罗华20岁创造了群论。赫姆霍兹26岁、迈耶28岁时提出了热力学第一定律。克劳修斯26岁时提出了热力学第二定律。开尔文24岁提出绝对温标，并预见到热力学第三定律的存在。麦克斯韦25岁对电磁理论作出重大改进，34岁建立起著名的电磁方程组。爱因斯坦26岁发表狭义相对论，提出光子说，36岁又发表广义相对论。历史上重大的科学发现，大都是年轻人作出的。他们虽然知识不如老年人丰富，但很少保守思想，最具创新精神。

社会科学和文艺创作与自然科学不完全相同。贝多芬在47岁的时候说：“现在，我知道如何搞创作了。”而且他在47岁之后的作品，确实比以前的作品更为出色。贝多芬和莎士比亚一样，都是在晚年才达到自己艺术创作的顶峰的。

然而，不管是贝多芬、莎士比亚还是其他杰出的文学艺术家和社会科学家，都不是40岁以后才开始建功立业的。他们中的大多数，在30岁

之前就已锋芒毕露，显现出耀眼的才华。莎士比亚写作《亨利四世》时只有28岁，发表《仲夏夜之梦》和《哈姆雷特》时也才36~37岁。贝多芬在30岁之前就已创作出优秀的作品，31岁时完成著名的《第一交响曲》，40岁出头完成8首交响曲和50多部其他作品。莫扎特是有名的神童。歌德25岁发表《少年维特之烦恼》，席勒21岁完成剧本《强盗》，24岁发表《阴谋与爱情》，雪莱21岁发表长诗《麦布女王》，狄更斯24岁完成《匹克威克外传》。泰戈尔15岁开始写剧本，普希金30岁之前完成了他一生中主要的诗歌创作。马克思30岁、恩格斯28岁时发表《共产党宣言》，毛泽东26岁主编《湘江评论》。亚历山大30岁左右就横扫南欧、北非和西亚，建立起横跨三洲的大帝国。诸葛亮27岁发表《隆中对》，提出三分天下的战略方针。打赢赤壁之战时，周瑜34岁，诸葛亮才28岁。

翻开历史的长卷，我们看到“自古英雄出少年”。青年人应该有志气，有抱负，完全不应在权威面前有自卑心理。应该像牛顿那样，努力站在巨人的肩上，让青春发出光辉。

青春的光辉，主要产生于勤奋而不是天才。爱因斯坦曾经说过：“在天才和勤奋之间，我毫不迟疑地选择勤奋。它几乎是世界上一切成就的催生婆。”《聊斋志异》的作者蒲松龄曾用下面的对联自勉自警：

有志者，事竟成，破釜沉舟，百二秦关终属楚；

苦心人，天不负，卧薪尝胆，三千越甲可吞吴。

清代诗人赵翼也在诗中写道：

江山代有才人出，各领风骚数百年。

轩中的这本书列举了大量科学发展的曲折经历，对增强读者的创新欲望，提高读者的创新能力，大有裨益。而且在这方面，轩中本人就作出了榜样，他像射门意识极强的足球运动员总想把球踢进球门一样，极想做一些创新工作。由于不断地拼搏努力，他在26岁的时候就完成了

《相对论通俗演义》一书。今年正当轩中的而立之年，他又呈现给读者这本优秀的作品。

我应轩中之邀，不揣冒昧给他的新书作序，好在他的书确实是本好书。

赵峥

2011年9月于北京半读斋

序2

最早听说张轩中是在百度相对论吧，后来连载的《相对论通俗演义》出版了纸质书还特地找来阅读，听说他还借此凝聚了一些人气，开办了一个网络学校，以爱因斯坦命名普及科学。此次他的新书《日出：量子力学与相对论》（下简称《日出》）即将出版，我很高兴地先阅读完此书，并向大家推荐。

开尔文在展望20世纪的物理学时提出了科技史上著名的两朵乌云：“动力理论肯定了热和光是运动的两种方式，现在，它的美丽而晴朗的天空却被两朵乌云笼罩了……第一朵乌云出现在光的波动理论上……第二朵乌云出现在关于能量均分的麦克斯韦-玻耳兹曼理论上……”他和当时绝大多数的物理学家一样，认为物理的大厦已经构建完成，所需要做的不过是对大厦粉刷粉饰罢了，未来的物理学者们大都可以失业了。他们所疑惑所轻视的乌云，导致了随后的科学革命，量子力学和相对论就此诞生。乌云散去，阳光普照，我们可以套用牛顿墓碑上的话来评价爱因斯坦等人：“上帝说要有光，于是他们诞生了！”

《日出》就是这样一本介绍量子力学和相对论发展历程的科普书籍，但是与其他科普作品相比又有着鲜明的特点特色。霍金曾在《时间简史》里打趣“听说作品里放上一个公式，就会吓跑一半潜在的读者”，于是他只在书中放上了著名的质能方程“ $E=mc^2$ ”，《日出》却毫不避讳地在书中直接引用各种公式和方程，显然没有语言比数学更优美、更精确、更打动人心的了，这是本书的第一个特点，保证了作为一部科普著作必须具备的科学性，而且具备了一丝大师的视角。（这是做科学研究和科学史研究的不同，后者并不需要你在科学研究方面成为大师，具备基本素养，仔细研究分析，自己未尝不能提出一套自圆其说的理论解释历史规律，本书作者就已经开始尝试这么做了。）但是如果全是公式，

就会和专业书一样枯燥乏味，于是作者力图用轻松幽默的语言来编排故事，这一点相信各位读者去看看作者之前的《相对论通俗演义》就能够体会。

作者在书中分条缕析地沿着历史的轨迹和物理上的联系，将量子力学和相对论的发展和进步一一展示在读者面前。当泡利、薛定谔、爱因斯坦等人灵光一闪的时候，作者运用了自己的想象将他们的灵感具像化了，比如薛定谔是和女友你依我依的时候，海森堡是望着大海等待日出的时候，虽然这些都是出于想象，但是又很好地结合了推导过程，读起来相当有趣，没有亲自学习一遍推导一番大概是很难写出来的。

这也是本书最大的不同：想象！如果其他大多数科普书是正史是《三国志》的话，这本书可能是野史、是笔记、是话本、是《三国演义》。读过《三国演义》的朋友都知道，罗贯中写三国有几个特点很鲜明：第一是大事件大历史方向保证正确但是在小细节小情节上多有妙笔，比如空城计；第二是人物个性鲜明，立场坚决，“红脸的关公，白脸的曹操”即是如此。《日出》也具备这两个特征：第一，物理史的大事件和逻辑推理内在联系没有问题，但是会用王小波式的幽默配上科学内涵深刻吐槽一些科学家和一些物理事件，或者如前所述加入一些花边新闻；第二，每个科学家的特点同样鲜明，比如“奥地利也是一个物理英雄辈出的国度。玻耳兹曼，以及本书后面将写到的薛定谔和泡利，他们便是来自维也纳的三剑客。依据他们的秉性，我们大概可以分别称呼他们三人为忧郁哥、多情哥和犀利哥”。三个人的性格不仅是心理学特征还带上了他们的物理学属性。

用这样的笔法写科普的不多，虽然将科普小说化也不算新鲜，比如科幻科普，国外有伽莫夫《物理世界漫游记》珠玉在前，其后层出不穷，从平面国续集《二维国内外：数字漫游奇历记》到最近的《三把锁的门：量子世界奇遇记》都堪称经典，国内也有《小灵通漫游记》，但之后就几乎消失不见。一些小说也开始逆向走学术路线，比如《剑桥金

庸武侠史》、《剑桥倚天屠龙史》。大概小说和科普自身都到了打破旧范式，寻求革命的时候了吧。这一点倒又和此书的科学革命遥相辉映，或许能从科学革命的结构中学到一点儿东西。

张轩中的科普之路一路走来，《日出》虽不能说是其集大成之作，但称为转型寻求变化之作并不为过，从《相对论通俗演义》不怎么像演义再到现在的《日出》，风格的变化显而易见。至于这种小说化的写法是否能受到大家的肯定还得大家捧场看了才知道。希望读者喜欢此书，能将历史上的物理和小说化的物理成功区分开来，甚至抛离小说化的背景仍能被物理本身打动，领略其中的天地大美。

戴一

2013年8月

序3

那天晚上，我与师父张轩中从北师大小西门的兰惠咖啡厅离开的时候已经是深夜，学校的小西门已经关上，我们只好横穿校园，从另一边的南门出去。在偌大的校道上不见一个人影，显得有点阴森森的——因为师父以前跟我说过，这北师大一带，所谓铁狮子坟与小西天，曾经是一些乱葬岗——但是远处闪着微弱光亮的路灯却似乎给了我们希望。

那时候是2011年，我只是一个普通的高中生，刚刚考完高考。同时我也是一个热爱物理的年轻人，并且有点理想主义。高中的某一天，我认识了师父，知道他写过一本物理方面的科普书，叫《相对论通俗演义》，还知道他见过霍金。在我彷徨的高中时代，他和朝愷（一位学物理的学长）把我带到了“理论物理”的大门前，并且给了我一把开启大门的钥匙——师父说，考完高考你如果愿意，可以来北京，我教你广义相对论。于是我一个人从广西来到北京，开始了我们的师徒缘分。

我们第一次见面是在北师大兰惠公寓一楼的大厅里，他给我带了一堆物理书作为见面礼，其中还有彭罗斯写的The Road To Reality，那是我第一次读到的英文物理书籍。我们一同吃完饭以后就到了他家里开始上第一节课。上课的头一句话，是师父问我是否能够背诵《孟子》“天将降大任于斯人也……”那段话，当时我背了出来。师父说，很好，你跟一般的高中生不一样，以后要牢记这段话。虽然我没有看出自己跟一般的高中生有什么不一样，但是我的课程就这样开始了。

那一个多月，我住在北师大附近小西天的一间潮湿的不足十平方米的地下室里，每天六点多起床到北师大的食堂吃早饭，然后到教七楼或者是物理楼的自习室上自习。到了下午五点多，到小西门的兰惠公寓等师父一起到新乐群食堂吃晚饭，再一起回到师父的住处，听师父讲课。我们课程的主题是“几何与广义相对论”，其实一开始主要是讲广义相对

论微分几何基础。他每天给我讲一个小时，总会留下几个问题算是我第二天的作业。虽然我高二的时候就学习过简单的微积分，物理的基础也并不算太差，但是因为师父留的问题开放性都比较大，所以第二天我都需要付出很大的努力才能完成。而每当我顺利完成题目的时候，他都非常高兴。有一次我突发奇想，用高中的排列组合知识给出了 N 维空间黎曼曲率张量分量独立个数公式的证明，他高兴得非要给我拍个录像发到网上去——很多时候，当我想起他感到欣慰而表现出来的那种高兴，我就想做出更多的事情让他满意，让他知道自己培养这个学生不是在浪费自己的时间。因此，那段时光我的生活特别充实，特别愉快。

偶尔，我们的“教室”也会是在北师大校园里，田径场上。我们一边散步一边讨论宇宙与时空，这个度规、那个曲率，嘴里蹦出这些个词语让路人咋舌，而我们毫不在意。甚至，在那个繁星闪烁的晚上，他忽然把我叫到兰惠咖啡厅，说请我喝咖啡。之后我就一边喝着咖啡，一边听他讲“参数共振”，感觉到从未有过的新鲜感。我想起在哪本科普书里，读到过量子力学诞生那个时代，物理学家泡利和海森堡也常常跟着他们的老师索墨菲在咖啡厅里讨论物理问题。而今我正在做着同样的事情，使得我心里有一种“跟上他们脚步”的错觉，顿时觉得前途一片光明。那次我们忘记了时间，一直谈到深夜，直到小西门都关闭了我们要穿过师大从南门出去。我们慢悠悠地走，在激烈的讨论之后，走出咖啡厅，享受无尽的夜。师父忽然说：“在我读本科的时候，也曾经在一个夜晚跟着我的老师漫步在这校园里，我们讨论假如我们不看月亮的时候，月亮到底还存在不存在这样的量子力学问题。如今我有了自己的学生，再过十年，你也会有自己的学生……”仿佛让我在无尽的夜色中看见光明的未来。

总之，在那个假期里，我学到了广义相对论的一些基础，对物理学也有了更深的认识。当时，本书还在创作中，而我有幸成为第一个读者。也因为师父的启蒙，使得我对物理学有了更深的兴趣。在我进入辽

宁科技大学物理系读书之后，受到老师们的赏识，学业得心应手，从根本上来讲也是由于那个假期打下的物理基础。因而在大一的上学期结束之后，我又再次回到北京，跟他学习量子力学。

那一次，我在北京待了整个寒假。这是我有生以来第一次没有回家过春节，而是在为自己的理想而奋斗。我们谈论氢原子的代数解法，我第一次知道龙格-楞次矢量是这么重要，我们在量子力学里面发现原来微不足道的扰动也足以影响整个系统——那么微不足道的我是不是也有一天能够有所成就？读者或许在这本书里能看到历史上的一些经历过不少挫折的少年，最后变成了阿贝尔，变成了薛定谔，变成了爱因斯坦。而我，在书中看到过自己的影子，也和书中的他们一样怀揣着成为一个物理学家的梦。

感谢师父，给了我圆梦的机会。

黄宇傲天
2013年8月

第一部分 量子力学

SUNRISE

A story of the Quantum theory
and Relativity



薛定谔和他的情人在阿尔卑斯山滑雪 绘画：田玉宽

1 孤儿：两个钻石王老五

(1)

1799年，法国的拿破仑（Napoléon Bonaparte, 1769—1821）正在准备一场旷日持久的战争，其第一阶段的目的是称霸欧洲。德国南部巴伐利亚州的一个12岁的少年刚刚失去双亲，他的苦难人生已经开始，他的父亲，是一个磨玻璃镜片的工匠，早早逝去。



弗朗禾费

这个少年就是弗朗禾费（Joseph von Fraunhofer, 1787—1826）。孤苦伶仃的弗朗禾费1799年8月23日离开了自己的故乡，来到慕尼黑一家玻璃作坊里开始了工作。他有一个严厉的师傅，除了玻璃方面的工作，弗朗禾费每天还要在师傅家里做家务。

劳动对于弗朗禾费来说也不算什么，真正令爱学习的他痛苦的是师傅不准他晚上点灯夜读。还禁止他去专门针对手工阶层的孩子所办的假日学校学习。

1801年7月21日，师傅家的房子突然倒塌，师母在这次事故中丧生。庆幸的是，弗朗禾费在事故发生4小时后在废墟中被发现，他几乎是毫发未损。这个大难不死的少年长大后成为一个著名物理学家，和牛顿一样终身未婚，甚至没有情人。他的私人生活也好像是一个谜，他公开的身份是一家光学仪器公司的经理。如果没有个人的努力，作为穷二代的他很有可能将继承他父亲的职业，成为一个磨镜片的工匠。但他奇迹般扼住了命运的咽喉。

1806年，一个叫乌茨施耐德的人开了一家光学仪器公司，他雇佣弗朗禾费为工厂工作。乌茨施耐德给弗朗禾费提供了凯斯特纳（Abraham Gottlieb Kastner）、克吕格尔（Georg Simon Kiug el）和普里斯特李（Joseph Priest ley）等人所著的光学课本。弗朗禾费加入工厂后的第一个任务就是打磨和抛光光学元件。

“上天给了我多少时间？”弗朗禾费心想，我命由我不由天，我得干点事情出来，“牛顿用棱镜做成了分光计，看到了七色阳光。那我应该怎么做才能站在牛顿这个巨人的肩膀之上？”

弗朗禾费慢慢积累自己的经验，苦苦思索。1813年，在一次光学实验中，弗朗禾费改装了经纬仪。他在原始经纬仪的物镜前加入了一块棱镜，棱镜会将白色光折射成组成它的彩色光和太阳光谱的黑色线，通过调整经纬仪与黑色线之间的交叉影线，就可以获得每条黑色线的非常精确的角度。弗朗禾费在暗室的百叶窗上开了一条狭小的缝隙，太阳光可以透过缝隙照射在棱镜上，透镜后面是经纬仪的小望远镜，这种仪器所展现的光谱与他在灯光光谱中看到的亮线不同的是，太阳光谱中出现了很多条黑线，这种暗线前后共有574条，他把最明显的26条从A到Z按顺序编排，这就是一直沿用至今的“弗朗禾费暗线”。长时间的孤独探索，已经使他成为一个典型的技术宅，他的实验能力得到了充分发挥，有了自己的光学实验室，发现太阳光谱中的这一年，他26岁。26岁，是判断一个物理学家能不能青史留名的关键年份，一般来说，26岁之前出名的

物理学家，多数有两把刷子，不可能靠忽悠欺世盗名，比如爱因斯坦就是26岁的时候发现了狭义相对论和布朗运动的统计秘密，从而奠定了自己的江湖地位。

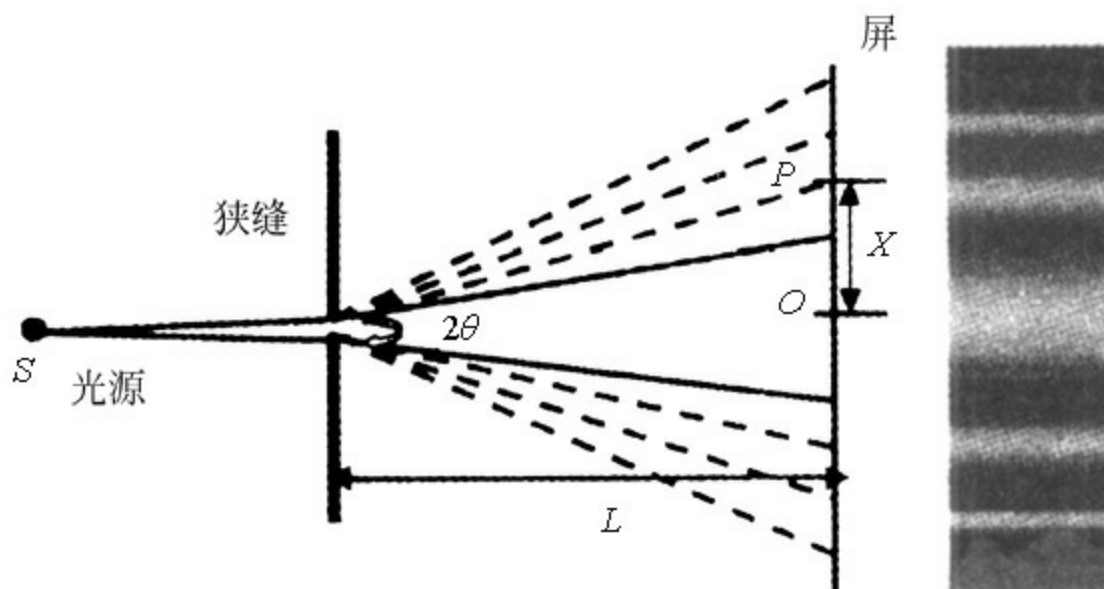
这些黑线，到底是怎么产生的呢？为什么会这样？

弗朗禾费黑暗线的波长主要有486nm，527nm，589nm，656nm，688nm，760nm，这些黑暗线其实是由于太阳大气对光谱的吸收，一般来说，弗朗禾费暗线宽度不到1nm——要探测到这些暗线，所需要的光谱仪器需要较高的分辨率。其实，后来人们可以看到，地球大气也能对太阳光形成吸收，一般的光谱仪能够探测到688nm和760nm这两个地球上的氧气对太阳光的吸收形成的弗朗禾费暗线。可惜的是，弗朗禾费黑暗线的重要性一直没有被物理学历史学家隆重指出，人们对弗朗禾费的宣传，也往往只停留在单缝衍射上，弗朗禾费单缝衍射是被所有的《光学》教材提及的。而弗朗禾费黑暗线实际上是必须用量子力学才能解释的现象——这是人类历史上第一次看到量子力学的背影。

从历史的角度来看，这个时候的德国著名数学家高斯（Johann Carl Friedrich Gauss，1777—1855）也还很年轻，他只比弗朗禾费大10岁。当他还是一个22岁的青年的时候，他完成了博士论文，证明了代数基本定理——这个定理说， n 次多项式方程总有 n 个复数根。在高中的时候，大家一定已经学过一元二次方程的求根公式，高斯证明的这个代数基本定理是一元二次方程的一个推广，该定理表明只要是多项式方程总是存在着解，你根本不需要担心方程在复数范围内会没有解。不过这个定理并没有说明如何寻找到那些解——如何寻找那些解答的程序成为历史悬念，后来被两个年轻小孩解决，一个是阿贝尔（N.H.Abel，1802—1829），一个是伽罗华（E.Galois，1811—1832）。高斯的这个定理在数学上叫做存在性的证明，虽然不具有操作性，但一般也是非常深刻的。高斯证明的这个定理不但使得他的博士学位含金量奇高，而且也可以用到以后出现的量子力学理论中，因为这个定理表明，一个 n 乘 n 的矩

阵，它的特征多项式就是一个 n 次多项式方程，这个方程有 n 个根。而这 n 个根，往往就是一个量子力学系统的离散特征值。

高斯是这个早期英雄时代的主旋律，他的思想博大精深，对数学和物理都有研究，因此他就好像是一条长江，是主流。而其他的英雄人物，则构成了他的支流。



高斯时代的另外一条支流，是同时代在法国的傅里叶（Jean Baptiste Joseph Fourier, 1768—1830）。

这个傅里叶并非那个空想社会主义学家傅里叶，而是一位著名的数学家，半个物理学大师。用傅里叶的话来讲，弗朗禾费实际上是发现了连续频谱之中缺失的几根小线条（这些离散的线条可以看成是离散谱线）。不过当时他们在两个敌对的国家，一个是法国人，一个是德国人，命运各自流转，天各一方，也没有国际交流的机会，因此，彼此不了解对方所做的学问。

傅里叶其实也研究太阳光，他是万般渴望温暖的人，更广义地说，他研究的是温度和热量……

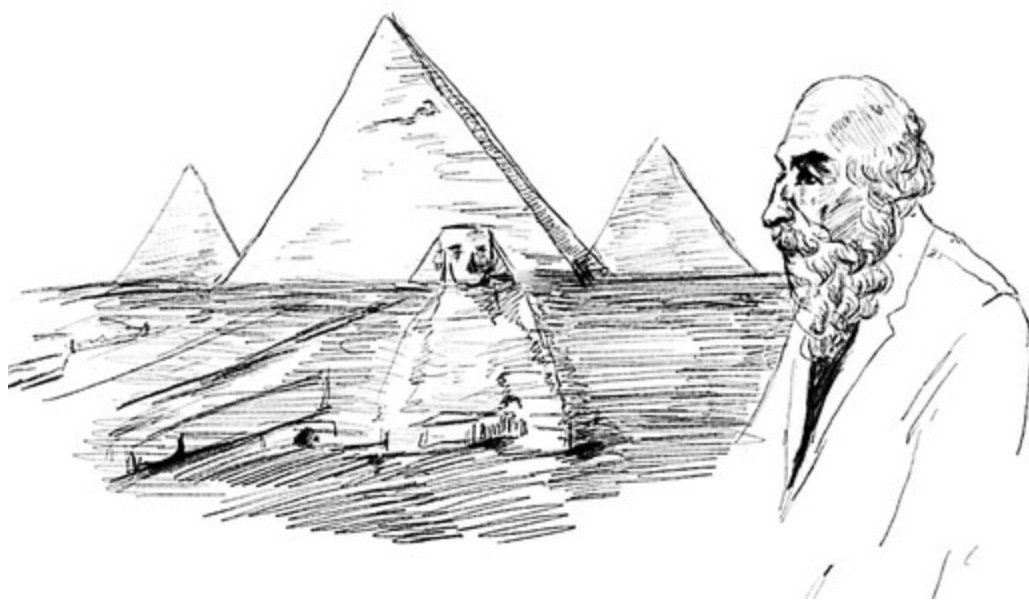
温度和热量，就在那个时代对数学家来说，还是很不容易搞清楚的几个物理概念！他们中的代表人物傅里叶，却要在这条路上走进一个神

秘莫测的世界.....

(2)

很多年以后，当傅里叶在官场几经宦海沉浮，在拿破仑与路易十八的城头变换大王旗下左右摇摆如墙头草，打击掉各种政治对手时，他是否还能回想起1789年跟着拿破仑远征埃及的那段青葱岁月.....

金字塔在沙漠里苍凉地矗立，冷峻而斑驳，漫天是飞舞的黄沙，让人睁不开眼睛，大漠的落日下，戈壁滩上有几只乌鸦在天际盘旋，似乎在寻找战死沙场的尸体.....一个30岁的青年用他那凝重的眼神深邃的双眸望着金字塔的狮身人面像出神，他的胡子已经很长，脸上写满风尘的疲惫。



金字塔前思索的科学家 绘画：贾宁

这个时候的傅里叶不是一个文艺青年，而是一位数学家、物理学家。傅里叶在很小的时候就失去了双亲，在贫穷中度过少年时代，在免学费的军校里读书成长。

拿破仑那次远征埃及，带着一个庞大的学者团，一共有165位学者

和专家，其中研究方向有地质、植物、动物、数学、物理等，各行各业，无不涵盖。拿破仑还册封自己为“法国科学院院长”，仿佛这次他不是来打仗，而是来搞学术研究的。傅里叶只不过是165名学者专家中的一位，他与著名数学家蒙日一起，主管军队的文化宣传工作。他本来也许不会在历史上留下名字，但这次埃及之行改变了他的命运：他得病了。埃及热带气候里的蠹虫和蚊子使得他得了一种很严重的病，这种病叫黏液水肿——一种让人总是感觉寒冷的疾病——也许是一种疟疾吧。

从法国到埃及的这次远征一事无成，但改变了傅里叶，他得了那个病以后，总觉得非常寒冷。于是，3年以后，当他从埃及回到法国，他被拿破仑任命为地方行政长官。在夏天他也要穿着厚厚的棉袄。因为实在是太冷了，冷得实在受不了的时候，他决定研究一下地球是如何获取热量的。如果他不叫傅里叶，学术界也许会忘记这个问题。当时傅里叶得出的结论是：尽管地球确实将大量的热量反射回太空，但大气层还是拦下了其中的一部分并将其重新反射回地球表面，所以，地球上的大气层有保温的效果。



傅里叶

傅里叶每天花大量的时间研究热量的传播问题，并提出了处理这种热量传递的数学工具：热传导方程。这个事情其实也不难，因为当时与拿破仑关系很不错的另外一位数学家拉普拉斯也提出了一个描写引力势能的偏微分方程，叫位势方程。傅里叶的热传导方程与位势方程的样子长得也差不多，前者是抛物型的方程，后者是椭圆型的方程。

他根据这个热传导的方程出版了一本很重要的书，堪比前辈牛顿的《原理》。这足够验证一句古诗“江山代有才人出，各领风骚数百年”。傅里叶的书影响了一代又一代人。书的名字就叫《热的解析原理》。在这本书中，他基本发明了一种新的方法，也就是频谱分析的方法。

傅里叶研究了热传导的问题以后，意犹未尽，顺便也研究了一下热辐射问题，比如说一个炉子所发生的热辐射到底满足什么物理规律。不过在这个热辐射问题上他完全没弄清楚是怎么回事——这就留给普朗克这些后来人来完成了。

在傅里叶的著作中他提到了一个很重要的概念，这个概念就是“傅里叶变换”。就好像白色的太阳光通过光栅或者棱镜可以色散为七色彩虹一样，傅里叶变换其实就是把任何一个信号转换成光谱的形式。光谱具有不同的频率，所谓频率，其实是单位时间内完成振动的次数，是描述振动物体往复运动频繁程度的量。傅里叶变换就是把振动随时间的复杂变化，转化为几个特征频率的线性组合。

同时在量子力学中，傅里叶变换也是把坐标表象和动量表象联系起来的工具，换句话说，傅里叶变换和矩阵一样是量子力学发展历史上一个绕不过去的存在。

写完《热的解析原理》这本奇书，傅里叶就不再做学问了，他开始更加努力地走行政路线。他在官场里浮沉，担任一个地区的行政长官，他的职责包括征税，征兵，执法，执行巴黎政府的其他命令，撰写政府工作报告。

(3)

弗朗禾费和傅里叶一样，都没有结婚。他后来成为一家光学仪器公司的总经理，销售他们公司自己做的光栅。虽然生活还算宁静，但太阳光里的离散暗线来源似乎是一个不小的难题，难住了弗朗禾费以及其他的很多人，这是新时代巨人扣开房门之前的沉闷脚步声。上天并没有给弗朗禾费更多的时间，作为光学家的弗朗禾费在有生之年还匆匆地干了不少事情，除了制造光谱仪，他留下了平行光线通过一个狭缝以后留下的衍射花纹——弗朗禾费衍射花纹。虽然早在中国的春秋战国时代，墨子他们已经发现，光线通过狭缝的时候，有的时候会出现小孔成像——在读初中物理的时候，我们就知道，这是“倒立的实像”。这事情只被做了一半，到了后来，弗朗禾费他们就注意到，如果这个狭缝开得很小很小，那么就可以看到衍射花纹，就像太阳光照到肥皂泡之上的那种五颜六色的花纹。

39岁那年，终生未婚的弗朗禾费离开了这个色彩斑斓的世界。他再也看不到这个世界的阳光和暗线。

当1807年已经凝固在时空长廊里，经常感觉寒冷的傅里叶正在油灯昏黄的灯光之下用鹅毛笔蘸着墨水写下热量传播的方程，并且他解出了这个微分方程——神奇之处在于，他可以用他的方法把微分方程变成简单的代数方程。

1807年的研究傅里叶翻开了历史的新篇章。而究竟什么是热量什么是热辐射这些问题还一直困扰着19世纪的人们。热量，热辐射，光？千头万绪涌上心头，而此时我们站在21世纪的山顶，回望来时路，看到的19世纪是山腰上的一座孤城。城上风光莺语乱，城下烟波春拍岸。风景很好，我们要慢慢欣赏那两位钻石王老五所留下的杰作。这里面有错综复杂的历史关系，也有思想上的激情澎湃和低潮期，读者们可都要做好了！

2 监狱里的群论

(1)

如果把1799年高斯写成他的博士论文那年作为英雄时代的开端的话，则之后的历史发展有一条清晰的潜规则，那就是数学往往在物理学之前先走一步。而在数学上，大约在1830年，伽罗华的稿件跟着傅里叶一起下葬以后，人们也跌跌撞撞地进入了群论的时代。

千年以前，人们开始研究代数方程，或者说是多项式方程。形如 $2^x + 3^x = 1$ ，这样的方程不算是多项式方程。多项式方程即在方程中的每一项都是由 x 的 n 次方项组成的。例如，一元二次方程就是一个多项式方程。数学天才高斯严格证明了代数基本定理，即 n 次代数方程 $f(x) = 0$ 成立，那么它必然存在 n 个复数根。但是真正在技术上求解 n 次代数方程，并不是很简单的一件事情。

数学史上一点一滴的进步，都凝结了前人的心血。即便历史容易被遗忘，一些英雄的名字依旧会被铭记。方程论上最早的英雄塔塔里亚（Nicolo Tartaglia），就在技术上解决了三次方程的求解问题。

塔塔里亚掌握了三次方程的解法，却没有发表。每天压在枕头底下暗爽，结果被人剽窃了。世道浇漓，剽窃的人反而成了当时该领域的学术带头人。塔塔里亚很是愤懑，1530年他约对方在米兰大教堂各出30道三次方程题目进行解题比赛。这场比赛观者千人，盛况空前，一时赚得娱乐圈和学术圈的头版头条。结果塔塔里亚大获全胜，对方一题未答。此事也成为剽窃史上的空前丑闻，让后人引以为戒。三次方程的解法尘埃落定，更多的人发起了对三次以上方程的挑战。

1824年，高斯48岁，功成名就，除了夫妻关系不怎么和谐，其他一切都还不错；伽罗华13岁，正在成长为一个愤怒儿童，他的父亲是拿破

仑的支持者，也卷进了法国大革命城头变换大王旗的洪流之中。与此同时，在北欧半岛的挪威还发生了一件默默无闻的小事，22岁的阿贝尔（Abel）自费出版了一个小册子，在这个小册子中他证明了对于 $n \geq 5$ 的 n 次代数方程，一般并不存在根式解。故事围绕着这些人展开。

（2）

阿贝尔是挪威的一个穷牧师的儿子，18岁那年，父亲去世，他差点因此上不起大学。当他还在中学时就开始着手探讨高次多项式方程的可解性问题——在初中的时候一般就会教授一元二次方程的求根公式与韦达定理，但更高次的方程的求根公式在中学里一般是不讲解的，在当时的欧洲也是这样的——阿贝尔后来上了大学，开始写更高深一些的文章，但命运不济，他写的关于椭圆函数的论文被巴黎科学院打入了冷宫。阿贝尔并没有放弃，又在不久以后发表论文证明了一般五次以上的代数方程，它们的根式解法是不存在的，只有某些特殊的五次以上的方程，可以用根式解法——这件事情后来在中国也引起一段故事，当时还在小杂货铺里算账的华罗庚就以一篇《论苏家驹之五次代数方程解法之不可能成立的理由》引起数学家熊庆来的注意，当时的华罗庚只有初中文凭，但从此进入清华大学，开始了波澜壮阔的数学人生——话说回来，当时阿贝尔这个文章没有引起别人的注意，阿贝尔于是急中生智，自费出书，并想借由这本书敲开高斯家的大门，于是，他把这本书邮寄给了身在德国的高斯。



在监狱里计算 绘画：张京

“如果高斯读了我的书，一定会惊艳于我的天才，他一定会约我，叫我去做他的帮手……”书邮寄出去以后，年轻的阿贝尔天天等着高斯的回信，他做着他的挪威梦。

然而高斯无情地粉碎了阿贝尔的梦。

没有回信。

寄出去的信好像是一块大石头沉没在无尽的深海……

急不可耐的阿贝尔亲自跑去哥廷根，想为自己的天才讨个说法。可是却没有见到高斯，伤心绝望之余，他跑去了柏林……

27岁的阿贝尔回到挪威，却得了肺结核，临死的时候贫病交加，除了一个女朋友愿意跟着他，其他几乎一无所有——这也不由得让人想起来那个在战后日本的废墟上自杀的天才数学家谷山丰。

天才生于寒冷，濒死之时，柏林大学邮寄出了给他的聘书，聘他去做教授。然而病重的阿贝尔并没有看到这封来信，一切都来得太晚了……

阿贝尔的理论对后世有着巨大的影响。他的研究成果在他死后轰动了世界，延续了3个世纪的五次方程难题终于获得了解决。然后问题却

似乎变得更加复杂了，究竟哪些方程可用根式解，哪些不能？这个更为深奥的问题浮出了历史的水面。

命运的齿轮似乎总在延续着惊人的巧合。阿贝尔去世的前一年，19岁的法国数学家伽罗华也写了一份论文交给法兰西巴黎科学院。他用一个新的方法阐述了能够根式求解的代数方程的条件，这是高斯博士论文的延伸，是一种对于求解方程可操作性的朦胧想法——基本意思就是寻找那个保持韦达定理的数学变化，重新排列根的位置与顺序，当保持韦达定理不变，这就好像一个大家族的男女在拍照合影，但保持各对夫妻关系家庭不变一样。

但是伽罗华的文章太过前卫，在别人看来有些南腔北调，不知所云。投稿2次，人家竟然把原稿给丢失了——前面讲到的傅里叶也得到过这个文章，但还没有仔细审稿就病死了。

很明显，伽罗华是另一位具有杰出才能的法国数学天才，他的出现在一定程度上也是法国大革命的成果。自从巴黎人民攻占巴士底狱，解救出八个政治犯后，法国就从封建王朝进入了资本主义社会，各种大小资本家们粉墨登场，最著名的就是以罗伯斯皮尔为领导的雅格宾派，在法国实行的红色恐怖主义的革命，杀了很多入.....这是伽罗华来到这个世界之前的20年，法国社会乱象丛生，思想界也异常活跃，在这个“旧制度与大革命”的夹缝中，出现了一个数学的奇葩。

1811年10月26日，伽罗华出生在法国巴黎一个小市镇上。这个时候的法国，已经进入了拿破仑时代，不过离他穷途末路也只有3年了。

1830年，21岁的刘维尔还是巴黎综合技术学校的学生的时候，他就写了一篇《热物理数学理论研究》的论文来申请当时巴黎科学院的最高数学奖，当时与他一起申请这个奖金的竞争者之一就是伽罗华。

很多年以后，以刘维尔名字命名的“斯图母-刘维尔方程”给出了后来所有量子力学系统中的能级。不说刘维尔，就说这个伽罗华从17岁就开始研究方程可解性问题，他实际上是创造性地提出了用群论的方法来

处理这类问题。所谓群论就是研究对称性的一门学问，其实很多时候人们早已经在利用对称性来处理数学和物理问题。比如在高中的时候，一些搞过物理竞赛的同学会被要求去计算一个正方体的处于体对角线上的两个顶点之间的电阻，这个正方体的每条边长的电阻是1欧姆。这个问题其实就是有对称性的，处于对称地位的点，它们一定具有相等的电势，所以可以合并成一个点，这样就可以很快求出这个系统的电阻是 $5/6$ 欧姆。

在本书的附录中，我们也提到一个用相似三角形的对称思想来得到狭义相对论的“时间膨胀公式”的工作——这个公式解释了所谓“天上一天，地上一年”的现象，这个工作同样是考虑了狭义相对论中关于参考系的对称性质而得到的精彩结论。（当然关于相对论的基本思想与内容，我们将在本书的第二部分加以介绍。）

前面已经说过，伽罗华也具有同样的思想，但是他更加深邃，他把方程的根看成是相互平等的对象，然后来研究这些根保持韦达定理运算的那些对称变化……最后，伽罗华得到了一个后世以他的名字命名的定理：一个多项式方程如果是可以用根式求解，那么这个方程的伽罗华群是可解群。所谓可解群也就是说，这个群模去它的子群后得到一个商群，这个商群总是阿贝尔群。

就这样，离散群论就因为处理代数方程的问题而诞生了。

(3)

那这与本章标题里的监狱又有什么关系呢？

话再说回到性格倔强的伽罗华，他比阿贝尔更加生不逢时，他3次把研究论文交法国科学院审查，都未能得到及时的肯定——直到1846年，他去世14年以后，才由刘维尔在自己控制的杂志上刊登了当年一起参加竞赛的这位伽罗华同学的文章——不仅如此，由于伽罗华热烈支持

和参与法国“七月革命”，在他进入巴黎高等师范学校的第一年就被开除学籍；当时的法国巴黎各派政治意见不合，习惯卸下门板，在街道上筑起街垒，互扔石头。这个时候拿破仑倒台，各种派系重新想轮流上台，开始又一场轮盘赌。伽罗华的父亲是追随拿破仑的，身为镇长的他由于被政治对手陷害，自杀了。

伽罗华也在这乱世的挣扎中因为共和党的一些思想与言论坐了几次监狱，所以，他的一些数学思想是在牢房里诞生的。

最后一次坐监狱，是漫长的8个月，在监狱里，他对群论的思想有了比较清晰的总结。在出狱后的一个月，21岁的伽罗华在一天晚上，为了一个女人，答应与人决斗。预测到可能遭遇不测的他在油灯下匆忙地写下了群论的纲领。这个纲领也算是一个遗言，在某个地方他写道：我的时间不多了……

第二天这位天才在决斗中牺牲。

1832年5月的这天。一轮血红的残阳挂在某一棵枯树的枝头。

整个世界都在哽咽。

阿贝尔和伽罗华皆在年轻之时就离开人世，然而他们对数学的影响却无比深远。他们对天才的年轻人有很好的示范作用，特引用歌词一首，以表哀思：

原谅话也不讲半句此刻生命在凝聚

过去你曾寻过某段失去了的声音

落日远去人祈望留住青春的一刹

风雨思念置身梦里总会有唏嘘

若果他朝此生不可与你哪管生命是无奈

过去也曾尽诉往日心里爱的声音

就像隔世人期望重拾当天的一切

此世短暂转身步进萧刹了的空间

青春请你归来再伴我一会

(4)

伽罗华的工作为群论奠定了基础——群论同量子力学和相对论有着密切的联系，这差不多是20世纪数学物理的主流之一。但19世纪的上半个世纪，群论的思想刚刚浮出水面，没有人可以预见到群论将来会和物理学有如此紧密的关系。

离散的群，就是说，群中的元素是有限个的，而连续的群，群元素则有无限多个。这需要另外一个数学家来发展它，当然是出于其他的目的。

挪威并不是一个泱泱大国，但它孕育了杰出的数学家阿贝尔。另一位大名鼎鼎的数学家索菲斯·李（Sophus Lie, 1842—1899）也出自挪威。其所发明的李群就是一种连续群，是研究相对论的基本数学工具之一。

索菲斯·李，他已去世八十年了。他的伟大工作完成于1880年，那时候他也还年轻。

1870年普法战争爆发——俾斯麦再度发力，他想要统一德国，除掉奥地利，在这个过程中他首先要打击强大的法国——当时年轻的挪威数学家索菲斯正在法国漂泊。他操着带普鲁士口音的法语。法国人认定他是普鲁士奸细，把他投入监狱。由于法国战败，形势一片混乱。当索菲斯的法国朋友最终找到关他的牢房并成功地使他获释时，那是1877年，他正静居囚笼，搞出了新的数学发现——他秉承了伽罗华在监狱里做学问的风格，但希望把群论的方法推广到求解微分方程，结果他也成功了。但索菲斯一直很失望，因他的工作没得到世俗的承认，他曾经为此而苦恼……等他后来稍有地位，他竭尽他的能力，整理出版了他的同胞阿贝尔的文集，他在监狱里的时候，第一个想到的人，是伽罗华，第二个想到的人，正是阿贝尔。

李群理论的奠基，也为后来的量子力学和相对论的发展奠定了数学上的基础，这好像是一股潜伏在大洋深处的潜流，在缓慢流动。

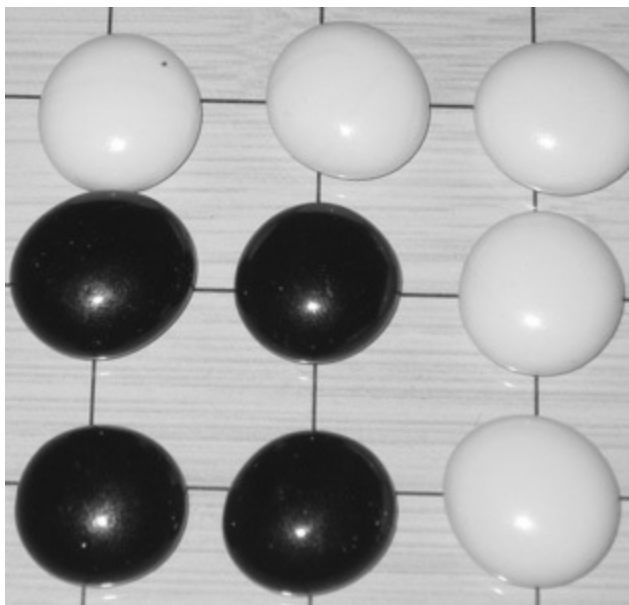
外一篇 伽罗华的死因考证，群论和组合数学

曾经也有人说他死于政治谋杀，总之愤怒青年伽罗华在政治上肯定只能算是一个无意义的炮灰，那位引起决斗的“风骚女人”，有人说是一位妓女，也有人说是一位政府密探、内奸，据最近的研究普遍认为她是伽罗华出狱后居住的旅店的医生的女儿——这个人物在历史上很隐秘，类似于我们后面要讲到的与薛定谔一起去滑雪的那个女子一样，姓名很难考证。伽罗华为了她主动挑起决斗。和伽罗华决斗的人是谁？伽罗华在遗书中说约他决斗的是两名“爱国者”。根据大仲马的回忆录，决斗者是当初被捕的19名军官之一德艾尔宾维尔。但是根据决斗几天后一家报纸的报道，与伽罗华决斗的是和他一起被捕的“人民之友社”成员、他的好友杜沙特雷。不管究竟是谁，这两人都是狂热的共和党人，也许不是政府安插在共和党人中的内奸——内奸名单曾在1848年被公开，这两人都不在其中。由于是朋友决斗，所以没有采取手枪对射的方式，而是采用“俄罗斯轮盘赌”，用枪口互相顶着对方开枪，其中只有一把枪装着子弹。因此伽罗华的决斗似乎不太可能是一场政治阴谋，而是由于一次恋爱事件。伽罗华之死被后人过度渲染，也形成了数学历史上少见的一段浪漫传奇，但无论如何，这死亡事件对伽罗华个人还是对科学发展而言，都是一场大悲剧。伽罗华如此轻生，也许与他因怀才不遇而厌世有关。他曾经两次报考被视为法国最高学府、大科学家云集的巴黎综合理工学院（该校当时的教授包括拉格朗日、拉普拉斯、傅里叶、泊松、科里奥利这些在科学史上声名显赫的大科学家），却都落榜，第一次是由于没有做好考试准备，第二次是由于顶撞考官。不得已才去上较差的巴黎高等师范学校。从此他就有了受迫害妄想。

再说回到群论。总之，所谓的群论，就是把很多操作看成是群里面的元素，比如下象棋，棋手每走一步棋，棋盘上的棋子的分布就会发生改变，那么棋手每走一步棋就是一个操作，这些操作在象棋的法则下构成了一个“群”（当然这只是一个比喻，不是严格的）；再比如，玩过魔方的人应该会有所体会，如果你把魔方的每一个面涂上数字，那么你每转动一次魔方，就是一个操作，这些操作很有可能会构成一个离散群。在量子力学中，人们为了研究晶体，处理那些有规则的多面体晶胞，也用到了同样的方法来研究它们的对称性。

群论的思想，起源于一些组合数学问题。人们在考虑一些数学问题甚至游戏的过程中，都可能会遇见对称性，这些对称性是很直观的。比如下图中的围棋放置办法，棋盘旋转 90° 以后是完全等价的。根据群论的办法，我们还可以求解以下问题。

问题：9个围棋，4黑5白，放在一个 3×3 的棋盘顶点，做任何刚性变换后，分布不重合的算一种，问移动棋子后，有几种不等价的放置方式？（如下图算是一种方式）

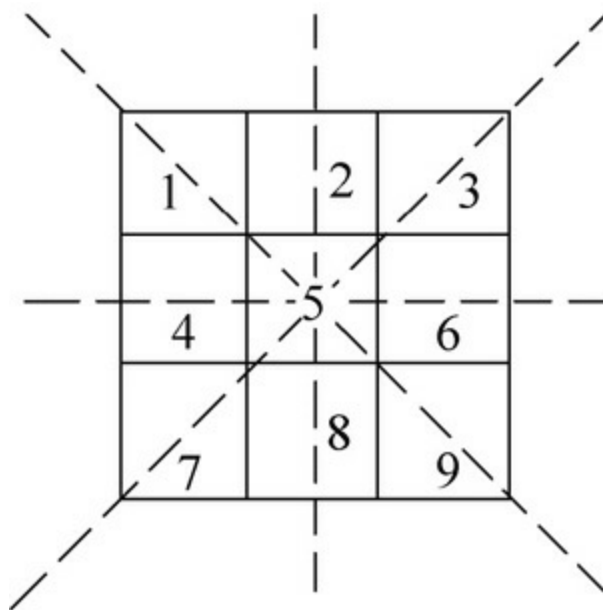


围棋

解答：以上问题可以转化为给3×3的格子染上颜色，4黑5白，做任何刚性变换后，分布不重合的算一种，有几种不等价的染色方式？

1	2	3
4	5	6
7	8	9

首先要分析一下它的对称性，这个正方形共有8个对称操作，3个旋转，4个反射，再加一个不操作（即恒等操作，旋转 0° ）。



四个反射操作

7	4	1
8	5	2
9	6	3

6	8	7
9	5	4
3	2	1

3	6	9
2	5	8
1	4	7

分别旋转90°，180°，270°的三个旋转对称操作

于是置换群G可以写成（群元e表示旋转0°，第2个群元（1793）（4862）（5）表示旋转90°的操作：把原来的1号变成了7号，把原来的7号变成了9号，把原来的9号变成了3号，把原来的3号变成了1号；把原来的4号变成了8号，把原来的8号变成了6号，把原来的6号变成了2号，把原来的2号变成了4号；保持5号的位置不变.....以此类推，就可以把整个群G用数学符号表达出来）。

$G = \{e, (1793)(4862)(5), (1397)(2684)(5), (19)(73)(46)(28)(5), (13)(46)(79)(2)(5)(8), (17)(28)(39)(4)(5)(6), (24)(73)(86)(1)(9)(5), (26)(19)(48)(3)(7)(5)\}$

我们根据这个群G的性质可以求出，对4黑5白的围棋来讲，一共有23种不等价的放置方式。

3 爱因斯坦与玻耳兹曼，马赫

(1)

新时代（也就是20世纪）渐渐地走近了，当镜头拉近，我们可以在巴黎街头车水马龙的流转中看到隐藏的时代背景：经过拿破仑的攻伐和普法战争的失利，一贯喜欢闹事的巴黎人民建立了昙花一现的巴黎公社，随即被扑灭……德国的一个铁血的人物，他就是大名鼎鼎的俾斯麦（Otto von Bismarck, 1815—1898）已经定型，他统一德意志并把德国带进了全面的军事化，这为第一次世界大战埋下了伏笔。这是一个渐变的乱糟糟的时代。这个时代的物理学也在发生静悄悄的变化，在这变化中有一个历史的横截面，那是1879年，这一年，爱因斯坦在德国的一个小城诞生。

爱因斯坦在未成天才之前，又是怎么样一个人呢？

1894年，爱因斯坦15岁，一个人留在德国读中学，他的父母一开始在德国，后来跑到意大利开设电器设备工厂，其规模都不大，订单少得可怜，与之竞争的著名企业有西门子公司等——当时西门子公司也没有现在那么大，但因为能够承接到慕尼黑等城市的电报市政工程，所以逐渐发家，做政府项目的赢利能力确实很不错的。

爱因斯坦并不喜欢做生意，他对家里的电器设备的原理掌握得还行。而且爱因斯坦是一个天生的自由主义者，他视任何纪律生活为仇，他不想服兵役——于是他开始考虑如何逃避兵役。他想翻过阿尔卑斯山马上离开德国去意大利跟父母团聚，他不想再做德国人。

“爱因斯坦，谁叫你不幸生在德国！”爱因斯坦在心里说。

1894年，未来的迷茫和混沌笼罩着年轻的小爱因斯坦，不知道何去何从。当时德国社会的军事化气味太浓了。国家领导人全着军装，整个

国家都像是穿着制服，让爱因斯坦渴望自由的心灵受到了莫名的压抑。2个月以后，爱因斯坦拿到一张医生开具的证明，说自己神经太衰弱，已经不能再读书了。于是，他被学校开除了。爱因斯坦成功地来到了意大利，度过了自由自在的一年。

这个时候爱因斯坦连中学毕业的文凭都没有。



绘画：张京

接着他来到了瑞士，去读高考复读班。他到的一个学校，名字叫做阿劳中学。他租住在别人家里。房东不是别人，正是阿劳中学的校长。校长是一名教授，名叫温特勒。住在温特勒家里，爱因斯坦跟打了鸡血似的天天都很兴奋，因为校长家有一个小女儿玛丽，长得很萌，虽然不懂物理学，但却是一个很好玩的可爱姑娘。爱因斯坦就这样不由自主地陷入了情网，玛丽成了爱因斯坦的初恋女友——这两人在墙根偷偷拥吻。

1896年的爱因斯坦还是一个小人物，他通过放弃德国国籍来打破他心中的旧社会，他成了一个无国籍的人士。这一年，他与初恋如胶似漆如火如荼。

总之对爱因斯坦来说这是甜蜜的一年！
他还不知道20年后会是什么命运等待着他！

(2)

当时的学术圈里，其实还是有几个高手的，这几个高手相互较量，他们想要得到一个关于原子的基础理论。这也暗示着，新时代应该有一套关于微观世界的完整的学术理论。



玻耳兹曼

这其中一个高手就是玻耳兹曼（Ludwig Edward Boltzmann 1844—1906），玻耳兹曼是奥地利的物理学家，专攻统计力学。统计物理好像物理学中的宏观经济学，研究的是大量原子的集体行为，所以很有一点难度。玻耳兹曼研究这个学问的时候有一个单纯的信仰，他相信原子的存在，虽然他从来没有观测到过原子——就是在现在这个时代，要想看到原子或者离子也是非常困难的，需要使用扫描隧道显微镜或者通过致冷型的荧光CCD摄像机才可以看到，这类典型的产品可以参考美国PI仪器公司的产品。

那是19世纪晚期的奥地利首都——维也纳，有着写下《一个陌生女人的来信》的茨威格和新年音乐会的金色大厅，这些也许代表了大部分中国人对这个城市的所有认识，这也是这个城市的灵魂之所在。其实单就音乐而言，奥地利还有天才钢琴家莫扎特。但是很多人也许并不知道，奥地利也是一个物理英雄辈出的国度。玻耳兹曼，以及本书后面将写到的薛定谔和泡利，他们便是来自维也纳的三剑客。依据他们的秉性，我们大概可以分别称呼他们三人为忧郁哥、多情哥和犀利哥。

忧郁哥坚定地相信原子的存在。

什么是原子呢？

古希腊时代的哲人认为原子是万物组成的最小单元，若用一把无比锋利的小刀去切割一块橡皮，小刀切下去直到橡皮不能再分，剩下的东西就是原子——这种最朴素的原子观念。而在古希腊时代的哲人的心目中，一个苹果和一块橡皮的本质区别仅仅在于相同的原子的不同排列——当时还有哲学家发现原子在空间能排列出来的正多面体只有5种，优美的空间排列结构非常稀少。现在看来这种关于原子的说法无比简陋且不够严谨，门捷列夫在19世纪中叶就开始得到元素周期表，说明元素其实是分门别类的，但古希腊哲学家们关于正多面体的结构研究则意味深长——这与三维欧氏空间中的离散群论有关。

到了后来，原子的观念有了一些变化，其被分为不同的品种，但不是因为它们构成的空间结构不同，而在于它们的化学性质不同（主要是起源于核外电子总数与最外层的电子数），它们共同构成了一张元素周期表。

但是，在19世纪晚期20世纪初这个新旧时代的分水岭，还是没有人能够通过各种物理实验或者化学实验直接用人的眼睛去看见原子。所以学术圈对玻耳兹曼所笃信的原子观念有一种天然的排斥。关于原子的概念，与之相关的典型物理问题是比热问题，在初中物理课本中也是最早讲的吸收热量与比热容的那个 $Q = CM\Delta T$ 这个公式，其中C是比热，M是

质量， ΔT 是温度的变化，这个是关于热量与温度之间的第一个典型公式。早在1819年，原是化学家的杜隆（P.L.Dulong，1785—1838）和物理学家珀替（A.T.Petit，1790—1820）进行了一系列测量物体的比热实验。他们选择的对象是各种固体，比如金子、银子、铜块和钻石。他们在室温下做了大量实验，积累了大量数据，对于许多物质总结出一条经验规律：“所有简单物体都精确地具有相同的比热，每摩尔（注意不是每千克！）的物质温度升高一度需要大约25焦耳（大约是6卡，因为1卡等于4.2焦耳）的热量，金子、银子和铜块差不多一样，但是，钻石不是这样的……”当时他们无法解释为什么钻石的原子比热和大部分贵金属的原子比热差异那么大。当然，那时候的钻石还不值钱，女人们在结婚的时候也不要求男方送上钻戒。原子比热问题的最终结果，需要量子力学完善以后才能得到彻底解决，这是物理上悬疑的一方面。

1863年，这个叫玻耳兹曼的大学生进入维也纳大学学习物理学和数学专业。物理学院的斯特藩（J.Stefan）是一个不错的物理学家。当时，斯特藩是物理学院院长、数学和物理学教授，在斯特藩的悉心指导下，玻耳兹曼学到了气体和辐射方面的基础知识，他们两人后来一起创建了黑体辐射理论的积分表达：斯特藩-玻耳兹曼定律。这个定律说的是，一个黑体的辐射功率与温度的四次方成正比——所谓黑体，就是那些与外界处在热平衡状态下发出辐射的物体，比如太阳、钢水，甚至以后我们会讲到的宇宙微波背景辐射，都是典型的黑体辐射。举一个例子，太阳的温度是由于太阳上的氢弹爆炸产生的，它的温度大概是5000℃。假设太阳上每秒钟内有2颗氢弹爆炸。那么，我们可以估计出，另外一个同样大小的恒星，如果其温度是10000℃，那么根据斯特藩-玻耳兹曼定律这个恒星上每秒钟差不多有 2^4 ，也就是每秒有16颗同样大小的氢弹发生爆炸。

玻耳兹曼大学毕业后成了斯特藩的助手，同时继续读博士，最后自然也成了教授。在得到斯特藩-玻耳兹曼定律以后，玻耳兹曼开始想建

立一个关于统计物理的完整的理论，他的理论模型中，最核心的观念就是原子的存在性。玻耳兹曼相信原子存在——这成为他基本的人生信仰。也恰恰是因为这个信仰，导致他一直都很忧郁。那时的他在大学里做物理教授，就如同每个江湖故事中的高手都有一个对手一样。与他PK的对手是马赫（Ernst Mach，1838—1916）——他的同事，此人被青年时代爱因斯坦视为偶像和精神导师。

马赫认为，原子既然肉眼看不见，也不能用实验检测出来，那么所谓原子就根本不存在。马赫的观点现在看来也是不无道理的。在现在的量子理论中，也非常重视可观察的物理量，不能被观测到的，那就不是物理量，如果是理论要求不得不存在的，那就只能称之为“鬼量”或者“鬼场”。单就马赫的观念——一个不能被探测到的东西，就是不存在的。这种说法无疑是非常符合现代的物理学理念的。但这里面其实有2个层次的问题，一种是原则上不可以被测量到，比如说一个量子的叠加态——以后我们会说的薛定谔的猫态——你一测量，就会被破坏。还有一种就是因为仪器与技术的局限，使得暂时不能被测量到的，比如说原子，三聚氰胺分子——这些东西随着技术的进步，通过特定科学仪器，最终是可以被测量到的。因此，马赫把这2个层次的问题打包到一起，抛给了玻耳兹曼。（作者注：可观测与不可观测还有第3层的概念，那与时空的结构有关，比如因为宇宙大爆炸以来的时间是有限的，光的速度也是有限的，所以我们只能看到有限的宇宙，看不见的那部分被称为不可观测的宇宙，在我们的视界之外。）

那为什么原子不能被光学显微镜看到呢？

因为可见光的波长大概是400~800nm。为什么有的光人的眼睛可以直接看到，有的光人的眼睛看不到？这是因为人类是在太阳系这个环境中进化出来的，太阳作为一个黑体辐射的光源，其表面温度是5800℃，而根据维恩位移定律则可以算出，太阳光辐射的最大辐射波长在500nm附近。所以人的眼睛对500nm为中间值的光波长范围最敏感，

肉眼可见的光波长范围为390~780nm。 $1\text{nm}=10^{-9}\text{m}$ ，而原子的大小大概是0.1nm，所以，光波长很容易绕过单个原子，于是，原子根本就不能被看到了。如果玻耳兹曼可以穿越到现代，那么现代的科学仪器能够让他直接看到原子，这种神奇的“放大镜”就是扫描隧道显微镜——利用的是量子力学中的所谓隧道效应原理。如果去北京北四环保福桥下的中科院物理所，在那里有扫描隧道显微镜，你也就可以一睹原子的芳容。

就这样，在不知不觉中，因为时代的局限，马赫和玻耳兹曼就原子的存在性问题掐了起来，耗尽了他们的青春。

(3)

马赫其实本质上是一位流体力学的专家，而且是最接近相对论启蒙思想的学者，但他不是一个好的量子论的启蒙思想家。马赫当时的脑子里最清楚的一个事情是关于声音的传播速度，声音传播的速度随着空气的密度与温度是可以变化的。换句话说，声音的速度是在空间点上的函数，而流体的速度也是空间点上的函数。所以，这2个函数的比率就定义为当时当地的马赫数。这是在任何飞机的制造和设计理论中，都是需要提到的一个基本概念。在一个空气流动的场中，当时当地的人测量到的声音的传播速率并不是一个普适的常数。比如说，在地球上，在上海的人在下午3点测量到的声音的速度，和在北京的人在下午4点测量到的声音的速度一定是不一样的。这思想其实已经很有点神似狭义相对论了——狭义相对论是后来爱因斯坦总结出来的一个物理理论，主要意思是说，在不同的地方，不同的人所经历的时间快慢是不一样的，时间流动的快慢是和这个人本身运动的速度快慢有关系的。爱因斯坦，也许正是站在马赫的肩膀上开始了他的狭义相对论创作的，很难讲爱因斯坦没有从马赫这里得到启发。

说回到原子论。

马赫的观点得到了另外一位化学大师奥斯特瓦尔德（德语：Wilhelm Ostwald；拉脱维亚语：Vilhelms Ostvalds, 1853—1932）的肯定，他是非常不相信原子存在的一位化学家。在现代人的眼光看，这是一个非常荒诞的事情，一位化学家不信仰原子论就如同一个现代的医生不相信蛋白质核酸的存在一样。不过那是在19世纪，奥斯特瓦尔德最核心的思想是这样的：“这个世界上，最基本的运动形式是能量。”这被称为“唯能论”，激烈对抗玻耳兹曼的“原子论”，奥斯特瓦尔德的唯能论虽然并没有太大的实际意义，但他当时在学术圈的地位也同样重要。

奥斯特瓦尔德也是著名教授（后来得到诺贝尔化学奖），可以说在当时的地位一点也不比玻耳兹曼低，所以他们两人也是针尖对麦芒，谁也说服不了谁。



绘画：张京

顺便再插一句，后来，奥斯特瓦尔德和爱因斯坦也曾有一段纠缠的经历，这事情算是爱因斯坦青年时成长的小插曲，那时候已经是1901年了。

当时刚从苏黎士联邦工业大学师范专业毕业的爱因斯坦没有找到工作，尝尽了生活的苦，世态的炎凉让他汗如雨下，他于是给学术大牛人奥斯特瓦尔德写信请求帮助。信中大意是：我拜读了您的大作，我对您的崇拜犹如滔滔江水绵绵不绝，又恰似黄河泛滥一发不可收拾……信的结尾爱因斯坦问奥斯特瓦尔德需要不需要一名实验助手，爱因斯坦表示自己也许可以去做实验助手。

但是奥斯特瓦尔德并没有给爱因斯坦回信，这让病急乱投医的爱因斯坦觉得很受伤，犹如被扇了一耳光一样脸上热辣辣地疼，可是家里的女朋友肚子里已经有了孩子，这日子还得坚强地过下去。所谓人穷志短，爱因斯坦又上杆子写了第二封信，说：“尊敬的教授，很抱歉，上次给您的那封信，我可能没有写清楚我的回信地址……今天我把回信的信封一起给您寄上，邮票我也已经贴好了，您只要直接把这个信封寄回来就好了。”

但是，奥斯特瓦尔德教授依旧没有理他，卑微的爱因斯坦又被抽了一耳光，眼冒金星，开始万念俱灰。这个时候，爱因斯坦的父亲也非常焦急。爱子心切的他为了让奥斯特瓦尔德能鼓励一下自己的儿子，做父亲的他也写了一封信，信里说：“尊敬的教授，很冒昧地给您写信……我的儿子爱因斯坦给你写了两封信，您都没有回……为了不使他过分伤心，请您回信鼓励一下我这个沮丧的儿子……万分感谢。”

这次老爹出马，按照道理应该有回音了吧，但是，奥斯特瓦尔德还是不为所动，拒不搭理这对可怜父子，冷漠得很。

言归正传，关于原子是否存在的争论已经白热化了。

玻耳兹曼对决马赫和奥斯特瓦尔德组合，1：2，明显力有不逮，单就辩论来说，以一对二的玻耳兹曼常常落得下风，因为当时科技水平有限。“无法直接观测到的原子”确实不容易被人们相信。历史上的这场“原子论”和“唯能论”因双方都缺乏决定性实验和证据而沦为“动嘴皮子”的科学争论，最后以玻耳兹曼的自杀离场而宣告终结。同事的抨击

和生活的压力最终压垮了玻耳兹曼，他在一个旅游胜地自缢身亡。

可惜的是，事情很讽刺，在玻耳兹曼自杀后的一年，皮兰就通过布朗运动实验确定了分子原子论。

忧郁哥玻耳兹曼厌倦了人生，他自杀了，但他的灵魂却壁立千仞。原子成为一个客观的实在，而我们的故事有了一个开始的基础。为了下文的行文流畅，我们不妨在这里对原子做一个基本的了解。原子是由电子和原子核组成的。原子的尺度大约是 10^{-10}m ，这也是电子的活动半径，因为原子核的尺寸非常小，在整个原子中“如同棒球场中央的一只蚂蚁”。原子的另外一个重要特征是原子核能够产生强大的电场把电子拉住，使得电子不能飞离原子核。这个电场比人类能制造的最强电场要强10000倍左右——可以估计出这个电场强度，只要你知道氢原子的电离能量是-13.6电子伏，而原子的半径是 10^{-10}m 。这个强电场的存在保证了电子总是在原子核周围运动。但是读者们请注意，到现在我们还没有讲到电子到底是如何运动的。这个是后话。

而自杀者玻耳兹曼所建立的统计力学成功地把微观世界的运动和宏观世界的现象联系起来，他能够处理了 10^{23} 个气体分子的集体运动并且用数学的形式给出了直观和具体的表达，以他的名字命名的常数直接把大量气体分子的平均能量同系统的温度联系起来。在他之前，人们不太搞得清楚微观的能量和温度的关系。虽然同时期也有人一直在思考比热的问题。比如说，同样在夏天，在太阳暴晒下的钢板和一杯水相比，钢板的温度升高的速度比水更快，但为什么会如此不同，这背后其实有量子力学的东西，但玻耳兹曼那时代，他可以模糊地认为，能量是随着自由度均匀分布的——这就好像在一个十字路口，从东南西北四个方向发生车祸的概率是一样的——，这就是经典统计里的能量均分定理。

玻耳兹曼的统计力学的能量均匀定理本质上是一个数学假设，这就是所谓等概率假设。这个假设说，在投掷骰子的时候，如果骰子里面是没有被灌铅的，所以，掷出1和6的概率，应该是一样的。等概率假设使

得这个世界变得比较简单，换句话说，玻耳兹曼摸不清楚一件宏观的事情它发生的各个微观因素所出现的可能性大小，只好认为这些潜在的可能性是平等的，它们有同样的概率发生。而熵的存在，尤其是熵与量子力学，熵与信息论，熵与引力场的对应关系，就像一面面照妖镜照出奥斯特瓦尔德的“唯能论”其实是一个很低级丑陋的理论系统。

玻耳兹曼他走了，但他曾经研究过2个问题，第一是原子的存在性，第二就是模模糊糊的等概率假设。



玻耳兹曼的墓碑

江湖上，有人开始读玻耳兹曼的书，并且他的一些想法也渐渐被世人理解，但到底熵与微观状态有什么关系，则需要后来者来阐释清楚。

总结量子力学发生之前的历史，也是物理巨大变革的前夕。之前已经发生的事情给这一理论的诞生奠定了历史基础：

1821年忙着在金属铂表面刻制光栅的弗朗禾费忙得汗流浹背，验证之前提到的他发现的太阳光原子吸收光谱中藏有的暗线。

1822年写完了《热的解析理论》一书的傅里叶正穿着棉衣，蜷缩在太阳底下。外界对他的书的热情程度没有让他感觉到一丝的温暖，甚至

有些寒冷。他用棉衣把自己裹得更加严实，默默地望着天空，思考太阳光的秘密。

1833年英伦的哈密顿（William Rowan Hamilton, 1805—1865）正在创造比牛顿力学更容易推论到量子力学的新力学。他的力学系统最重要的特点是用动量与坐标来实现对运动的描述。而抛弃了速度的概念。

1870年挪威的数学家索菲斯也发展出来了李群（Lie group）——他一开始是为了拿这个东西去解决微分方程，因为在代数方程领域是存在求根公式和伽罗华的理论的，而微分方程领域还缺少类似的方法。

这些人和他们的研究正是19世纪的物理学大拼图的若干小碎片，这些碎片其实并不能完整地拼成一副名画，因为里面还存在一个关键的片段还没有完全搞清楚，这个核心的概念就是原子——到底什么是原子？它真的存在吗？

4 写清楚熵公式的人

(1)

1900年的车轮朝我们驶来。一个波澜壮阔的大时代就要拉开帷幕。

1900年是一个历史的分水岭，在这之前被称为19世纪，有人总结了19世纪科学的三大发现：热力学第一定律（能量守恒定律），进化论和细胞学说。这总结虽然高屋建瓴，但其提倡者恩格斯后来写了一本所谓《自然辩证法》，有人拿去给爱因斯坦看，结果被爱因斯坦批判了一番，因此结下梁子，引起后来社会主义阵营对爱因斯坦相对论与爱因斯坦本人的批判浪潮。从物理学的角度来说，19世纪其实还有一个重要的发现，那就是发现了热力学第二定律。这个定律的发现和瓦特发明的蒸汽机的大规模工业应用有关系，早在19世纪30年代，德国就有一个叫卡诺（Sadi Carnot）的工程师，他很想知道什么样子的热机，能达到最大的效率，通俗地说，就是这个机器，吃得最少，干得最多——在初中的物理书里就讲过，机械效率等于有用功除以总功。但有没有什么热机，它能达到100%的效率呢？——通俗地说就是公司招聘一个员工，这个员工100%的把时间投入到工作之中，这样的员工存在吗？



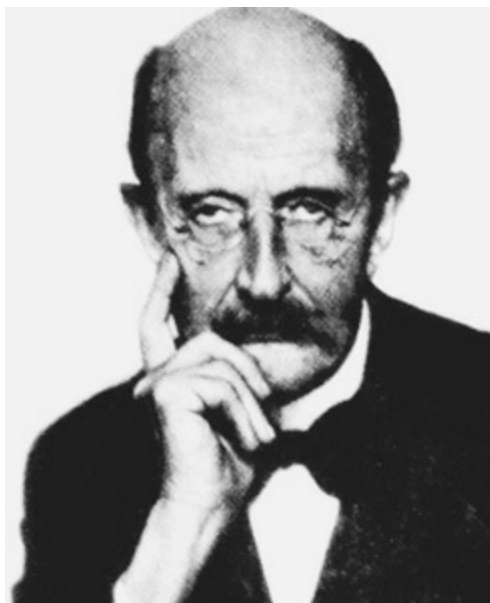
绘画：张京

卡诺开始了他基于热质学说的不清不白的研究，因为搞不清楚热量到底是什么，所以他的研究并不符合现代科学的正统，但他最后研究出一个所谓卡诺循环，按照那个循环工作的热机，其效率最高。

别的研究者前赴后继，到了1880年，基本已经有了一个较为完善的热力学第二定律的双重表述：不可能存在100%效率的热机。

在1900年之前，一个集热力学理论之大成的人要出现了，这一个物理新秀，他把前人的这些思想总结出来，清晰地给出了一个数学公式， $S = K \ln W$ 。

这个德国的年轻物理学家，就是普朗克（Max Karl Ernst Ludwig Planck, 1858—1947）。公式中S表示熵，而W是微观状态的数目，有时候也被称为相空间的体积。



普朗克

普朗克所做的这个数学公式，其实全部是玻耳兹曼的思想，因此这个时候他还没有多少创新精神。普朗克其实一辈子都算是一个保守的革命者，他后来鬼使神差地把物理学的大船开进量子力学的港湾，做了一次英雄船长。这是后话。

玻耳兹曼计算微观状态数的核心思想，基本上就是组合数学的方法，我们可以举一个例子：把3个水果（分别是桃子、苹果和菠萝）放进2个抽屉（一个抽屉在写字台上，一个抽屉在床头柜上）里，有几种放置的办法？

因为每个水果的去向有2种可能，因此一个一个地分步骤来放水果，一共分3个步骤，根据乘法原理，一种有8种可能，也就是对应8个安放水果的微观状态。这个时候，我们可以套用普朗克所做的那个公式，系统的熵正比于 $\ln 8$ ——换句话说，如果你是一个男生，和女朋友玩这个游戏，把你的眼睛蒙上，你女朋友把3个水果随机放在2个抽屉里，叫你去猜，对你来说，这个系统的熵正比于 $\ln 8$ ，你猜对的概率是 $1/8$ 。

玻耳兹曼当时也是这样思考问题的，当时他继承了英国物理学家麦

克斯韦的分子运动论的一些思想，考虑把一个箱子用一个隔板分成大小相等的两半。从统计力学的角度来说，气体分子不可能全部处于盒子的右边，而使得左边保持真空，为什么呢？因为这样的事情，发生的概率实在太小。从物理上来说，如果这样的事情可以发生，就会引起右边箱子的温度高于左边箱子的温度。我们可以计算出左边箱子有 n 个分子的微观状态，你可以先从1000个里挑选出 n 个分子的组合数，这个数字的对数，就是有 n 个气体分子位于右边箱子里的可能性大小，这个分布所对应的熵是可以计算的，它的数值大小，给了我们一个判断依据，可以判断一些事情可能不可能发生。在这个问题中，熵的最大数值在 $n=500$ 的时候取到。

(2)

普朗克写出了熵公式以后，当时物理学界开始流传两个未解之谜，一个是麦克耳孙（Albert Abraham Michelson）和莫雷（Edward Morley，1838—1923）在1880年就开始去测量地球在太空中运动的绝对速度，他们希望这个速度是相对于一个绝对静止的背景产生的，他们也天然地希望在地球上，迎着太阳光走和背着太阳光走，应该测量到不同的光的速度，但是，在实验中，他们似乎没有测量到光速的差异。这就好像是说，两辆相向而行的火车，和同向而行的火车，其速度之差是一样的，这让人很奇怪。当时因为没有电子技术，所以要处理微小的差异并不容易，但麦克耳孙和莫雷用光学的干涉仪做了这个实验，并且留下了第一个疑问，那就是一个绝对静止的背景是不是真的存在？运动到底是发生在什么舞台背景之上？

另外一个未解之谜则来自钢铁工业，德国当时处于工业化浪潮之中，产业在升级。大炼钢铁的过程中很自然的产生了一个技术问题，那就是如何测量铁水的温度？

铁水的温度大致上与它的颜色有关系，而颜色是由铁水发出的光波的波长决定的。当时在德国有一个叫维恩（Wilhelm Carl Werner Otto Fritz Franz Wien, 1864—1928）的物理学家，得到了一个经验公式，他认为，决定铁水颜色的最主要的光波长和铁水的温度是成反比的，这被称为维恩位移定理，实际上这个定理是后来普朗克所发现的那条曲线的微分表达式——也就是曲线的极值所在点。这个定理非常实用，虽然维恩也不知道为什么会存在这样的定理，总的来说，这个定理就好像汽车的牵引力和速度之乘积是汽车的额定功率是一个常数那样，这背后有很复杂的多冲程的汽车发动机的工作原理存在，但当时是搞不太清楚。高温铁水发出的光，被称为黑体辐射，这个辐射有一个总的功率，当时玻耳兹曼和他的老师已经得到了这个总功率，是与温度的四次方成正比的。这个叫做斯忒番——玻耳兹曼定理，也就是后来普朗克所发现的那条曲线的积分表达式。这个温度四次方成正比的定理，和后来德拜发现的晶体热容的温度的三次方定理一起，构成了量子力学历史上的两个与温度简单关于的优美定律。

普朗克是以熵公式起家的，所以，他思考物理的时候，与别人不太一样，他不但从能量的角度思考黑体辐射曲线之谜，而且还从熵的角度加以分析。眼光不同，看到的物理自然也就不同。

对于黑体辐射来说，要解决的问题是，黑体辐射的功率谱密度到底是满足什么方程。当时人们已经知道，辐射可以作为气体，也有压强，也有熵和内能，从各个角度都可以证明辐射气体的能量密度和压强成正比，只差一个常数 $1/3$ 。所以，光会产生所谓光压，一束光打在电风扇的叶片上，电扇叶会旋转。但是，辐射气体的熵和能量密度到底有什么关系呢？不同的学术流派得到不同的结论。

一种流派得到的熵和能量密度的微分与温度负的一次方成正比。而另外一个流派得到的熵和能量密度的微分与温度的负二次方成正比。这两个结果都不完全符合实验，实验家鲁本斯告诉普朗克，这两个流派的

理论一个在长波处与实验相符合，一个在短波处与实验相符合。普朗克听到这个消息，他决定做一个简单的裁缝工作，把那两个一长一短的裤管做成一条裤子。

办法非常简单，用通分的办法就可以把两个式子整合起来，引进一个待定系数就可以。这个方法在实验数据的处理中非常常见，就是“内插法”。这个办法就好像要求一个班级学生的身高平均数值，我们可以先求出男生的平均身高，再求出女生的平均身高，然后再整合起来，得到一个新的平均数值。这对普朗克来说，一点也不难。

他把公式写出来，就是一个简单的微分方程，一积分，就得到了后来被当成普朗克黑体辐射曲线的那个方程。这个方程的曲线画出来很像一个少女的乳峰。普朗克也是莫名其妙，不知道为什么会这样。

“反正我就把结果发表出来，让那些实验物理学家去看看符合不符合他们的数据吧。”

于是，普朗克发表了他的乳峰曲线方程。这一天是1900年的10月25日，发表以后，他也知道这完全不像大物理学家的作品，因为他说不清楚这背后的物理道理。

5 富二代德布罗意：凌晨旧戏

(1)

1909年，春天桃红柳绿，巴黎的塞纳河河水盈荡，岸边的法国梧桐郁郁葱葱，曼妙得像一个个风姿绰约的少妇，香榭丽舍大街和凯旋门金光闪闪，在浪漫之都巴黎这个温柔乡里德布罗意过得很开心。作为一个年轻贵族他刚从巴黎大学历史系毕业，他的父亲正是法国的一个伯爵，并且正是一位当权的内阁部长。



德布罗意，1892—1987

在一个历史学家不可以查证的夜晚，德布罗意突然意识到懂物理学的人才是真正的贵族。他哥哥是做X射线的，是一位物理学家，德布罗意觉得哥哥是幸福的。家里有金山银山，他实在很空虚，于是，德布罗意开始学起了物理，拿起了他生命中的天书——刚开始，物理书中每一个字都有豆腐干那么大。

“反正我就是来玩票的，但我最好也要搞个物理学的博士学位。”富

二代德布罗意心想。但是，不巧的是，1914年，第一次世界大战的硝烟冉冉升起，战争像一个美女一样被这个富二代邂逅。不期而遇以后，很多人像空气中的灰尘一样地卷了进来，比如天文学家史瓦西和物理教师薛定谔。这个时候薛定谔还完全没有量子力学的概念，他只隐约地知道丹麦哥本哈根有一个叫玻尔的年轻人建立了一套电子绕原子作圆周运动的轨道模型，来说明原子发出的光谱线。但其背后的物理稀里糊涂，显得毫无章法。

这个时候，德布罗意也入伍了，作为一个富二代，他当然不需要去前线冲锋陷阵当炮灰，他被分派了一个无线电技术人员的工作。在这个过程中，德布罗意对电磁波有了很深的了解——天线尺寸应大于电磁波波长的十分之一，这是微波通信里最重要的内容之一（一般认为天线长度是接收和发射波长的 $1/4$ 时，转换效率最高），正因为如此，如果用人说话的频率（10000Hz）去设计天线，你会发现天线的尺寸可能要几万里。但一般的手机天线是非常短小的，不会超过10厘米，那这是为什么呢？其实人说话的声音本身的频率转变成电流信号的时候，这个电频率会很高，可以达到MHz甚至GHz的级别。在电路中可以改变信号的频率，这种改变的方法称为“调制”，调制以后，电磁波的波长差不多和手机的尺寸是一样的。

因此，德布罗意对电磁波有了一定的了解，虽然他是读文科出身，但不知不觉地在无线电的技术岗位上知道了一个重要的物理事实——电磁波是有波长和频率的——对于这一点他非常清楚——频率和波长的乘积是光的速度。

“原来是这样啊，物理好简单啊，波长和频率原来是有关系的！”德布罗意心中窃喜。

谁也没有想到，等战争一结束，这个纨绔子弟就要在笔尖发现另外一种波，这种波非常之特别，称为“物质波”。这种波出现以后，薛定谔才发现了他的波动方程。

(2)

战争终于结束了，这时候是1919年，中国这边是北洋军阀统治时期，当时要开巴黎和会，商量怎么处理战后事宜，中国代表团准备签署丧权辱国的卖国条约，这引起中国大学生们的不满，开始在北京游行，史称五四运动。中国开始思想启蒙运动，两年后，就是中国的富二代徐志摩在诗刊上发表介绍爱因斯坦相对论的文章。

德布罗意也在战争结束后，决定去读博士，他找了一个法国巴黎物理圈的明星博导，名叫朗之万（Paul Langevin）。



朗之万

朗之万在收德布罗意做徒弟之前就已经是一位大师，他因为当时实在很出名，于是坊间开始流传一些桃色绯闻，主要声称他与刚刚寡居的另一位女物理学家居里夫人有些关系，一时间轰动全巴黎。因为居里夫人是研究原子射线的专家，得过诺贝尔物理学奖，是一个家喻户晓的人物，所以，他们两人要是有暧昧关系，这就好像一枚深水炸弹，在巴黎炸开来。巴黎本来是浪漫之都，但物理学家一般还是给世人以清教徒的

形象，现在这奸情确实是震撼人心。

早在1906年朗之万就研究很复杂的问题，比如布朗运动。历史是一个任人装扮的小姑娘，当装扮这个小姑娘的时候，我们也许发现，1905年以后，凡是研究过布朗运动的人，多数成长为大师，比如爱因斯坦，比如伊藤（Ito）。朗之万做的事情非常简单，他模仿了牛顿第二运动定律，把作布朗运动的花粉粒子的运动方程用牛顿第二定律 $F=Ma$ 的形式写了出来。不过，因为粒子是受到随机的不可预测的撞击力的影响的，所以，朗之万的方程里的力是“随机力”。随机力 $F(t)$ 因为是非常随机的，所以不可能写出它随时间变化的解析表达式来，但是，你可以用深邃的目光来看随机力——你可以写出随机力的“功率谱”。换句话说，你可以把随机力的作用当作是一个非常复杂的发光过程，但你可以写出这个发光过程的“光谱”来。

朗之万举重若轻义无反顾，用随机力的方法得到了布朗运动的规律。他的随机力实际上还是经典随机现象。

随机现象有很多复杂的起源，为了行文方便，我们把随机现象分成两类：

1. 经典因果性的随机现象
2. 量子统计性的随机现象

对量子力学来说，“随机现象”是非常重要的，如果你可以预测一个现象必然发生，那么这个现象对你来说，其实不包含任何信息。可以举一个男性读者们应该深有体会的例子，凡是能被观测到的现象，不一定是包含对你有用的信息——很简单，一位女士只穿着乳罩出现在你面前，如果这个现象出了你的意料，那么这个现象对你来说是包含信息的，反过来，如果你并不感觉这个现象振奋人心，这说明该现象其实不包含什么信息。物理学家只把那些“出人意料”的现象称为“随机现象”。

粒子的布朗运动看上去很复杂，本来是一个随机现象——你无法预测粒子在下一秒到底出现在哪里。但它依然是满足经典因果性的一个现

象——牛顿力学就能解释它。所以，朗之万方程里并没有德布罗意想要的东西。

德布罗意想要的东西，是什么呢？

这个时候，他也30岁了，时间紧迫，他还要赶工期写出一个博士论文呢。

(3)

1924年，德布罗意躺在床上，看着天花板上悬挂的吊灯出神，他若有所思，想确定一个博士论文的题目。德布罗意思绪翩翩起舞，心想如果这盏吊灯倾泻下来，我想我不会再存在，但博士论文总得先写出一篇来.....

时间已经到了凌晨，夜已深，对面的古铜色的闹钟在不停地转着秒针，德布罗意觉得有点莫名的兴奋与紧张，博士论文到底写什么好呢？毕竟是半路出家，自己也隐约觉得自己的物理，实在是有点糟糕的。

外面街道上有人在吵架，德布罗意打开窗户，看见街面上有2个醉汉正在围绕着电线杆作布朗运动，边转动边骂人，满嘴污言秽语。远处还有一个穿着鹅毛黄大衣的女人，在一家咖啡店门口等待着什么。天色已经那么晚，周遭是无边的寂寞笼罩下来，德布罗意想下楼出去走走，吹一吹冷风，也许脑子能够清醒起来.....

突然，灵机一动，俗话说得很好，“天下文章一大抄”。

他觉得自己应该抄袭模仿一个人的学问。

但万事开头难，重要的是先确定到底是抄袭模仿谁？

抄自己的导师吗？看来不行，兔子不吃窝边草。抄哥本哈根的那个物理学家玻尔的吗？也许可以，不过玻尔的原子模型非常僵硬，电子轨道是强行规定的圆轨道——这是谁都没有见过的轨道，跟意淫也差不多了。玻尔有些东西很费解，德布罗意对对应原理，似乎有点搞不大清

楚。那抄什么人？伦琴？对于X射线德布罗意相对比较理解，但这里面没有多少理论物理方面的内容，X射线衍射与X射线荧光，这些属于自己哥哥搞的东西，自己如果上去写这个东西，很难不被别人说不是自家哥哥帮的忙，再说，现在大街上很多妇女都已经理解X射线，妇孺皆知的东西，德布罗意作为一个贵族，是不屑一写的。

考虑了一圈引用借鉴的对象，德国物理学家爱因斯坦的形象渐渐浮现在德布罗意的脑海里。眉头一皱，计上心来，德布罗意脸上露出了诡异的微笑。就这样决定了，抄袭模仿爱因斯坦——因为爱因斯坦说，无质量的光子具有波粒二象性，那么，现在你们逼我出绝招，我德布罗意认为，有质量的电子，也具有波粒二象性.....

德布罗意在那个凌晨决定推广爱因斯坦的“光子波粒二象性”，变成“电子波粒二象性”。

但先要在细节上运筹帷幄一下。德布罗意拿出草稿纸写起来。

玻尔的“太阳系原子模型”让电子强行满足一个量子化条件。这样做电子其实是不自由的，玻尔把自己当上帝了。德布罗意心想，把电子解放出来，让它们自己做主吧。

如何赋予电子一个基本的性质，让它们自觉地表现出量子化现象呢？德布罗意希望把轨道和驻波联系起来。因为早已经决定抄袭模仿爱因斯坦了，跟着爱因斯坦走，所以干起来真是畅快淋漓，简直可以一气呵成。

根据爱因斯坦质能方程，如果电子有质量 m ，那么它一定有一个内禀的能量 $E=mc^2$ 。好，爱因斯坦和玻尔对光子使用了如下关系 $E=h\nu$ 。推广到电子，这个关系还成立，电子也具有一个内禀的频率 ν 。把两者联系起来吧，因为 $E=mc^2=h\nu$ ，所以他得到了第一个关系式

$$\nu = mc^2/h。$$

有了频率了，很好，那怎么搞出一个波长来呢？德布罗意想起了自

已在当无线电技术军官时候所知道的一个真理，毕竟电磁波是既有频率又有波长的，那我要电子也有波长！写到这里，德布罗意脑子已经很乱，心想，死马当活马医了，继续出招吧，让参考系变换起来！！

德布罗意心想，在狭义相对论的参考系变换下，波动形式是怎么变化的？基于这个思路德布罗意三下五除二就把电子的波长和动量联系起来了。得到了第二个关系式

$$\lambda = h/p$$

在草稿纸上得到上述两个关系式以后，他的博士论文其实已经写完了。蓝色的钢笔墨水还没有干透，但他的内心已经湿透，好像是经历了一场滂沱大雨，筋疲力尽。

这种有质量的粒子所附带的具有频率和波长的波被称为“物质波”——不携带能量，但有波长。“物质波”总的意思是说，有质量的物体，总伴随着这个波。

夜真的已经很深了，深夜的寂寞让德布罗意觉得自己的文章如此美丽。简直太美丽了。他喃喃地说了一句“怪你过分美丽”。说完就起身把手稿用火点着了。火光格外温暖，似乎是思想在燃烧。纸张斑驳地在火焰里舞蹈，不一会儿房间里充满了一股灰烬特有的焦香味。

德布罗意烧了第一份草稿，他知道这些东西已经刻在自己的灵魂里，不可能被忘记。

他决定好好睡一觉，第二天把整个过程清楚地写下来。

不久他的博士论文正式在江湖上出现，各大门派皆为之震动，他显然掀起了轩然大波，因为他的文科出身，再加上他富二代的身份，以及他构造的那种具有质量粒子都有的波长公式，让人匪夷所思。

众教授看完他的文章感觉脊背上有一股寒意，有的感觉自己是吃了一只苍蝇，有的则像是看到一则很搞笑的冷笑话，纷纷惊呼：“这富二代做物理果然有点胡闹啊。”

(4)

当初的朗之万是不是碍于情面想帮德布罗意混得一个博士学位已不得而知，但他读到德布罗意出品的博士论文以后，确实也被雷到了，他实在没有办法不给这个部长的儿子一个面子，于是，他把这篇博士论文邮寄给了在柏林的爱因斯坦。

信的内容大致如下：

尊敬的爱因斯坦阁下：

在我这里有一位研究生，已经攻读了五年的博士学位，如今即将毕业，在他提交的毕业论文中有一些新的想法.....

请对他的论文作出您的评价。

另外顺便向您提及，该研究生的父亲是弊国的一位伯爵，内阁的某某部长，若您.....将来您来法国定会受到隆重的接待.....

爱因斯坦收到信后，马上读懂了这里面有两个意思，其中一个是一人是富二代，法国高官的儿子，不好得罪，第二个意思是，这篇文章的思想完全模仿自己当年对光子的波粒二象性的说法，很是面熟啊。

爱因斯坦于是回信说，此博士论文还是很有创新思想的。

此时的爱因斯坦虽不属于任何名门望派，却已独步于江湖，颇有威望。有了爱因斯坦的这一封信，评审委员会的几位教授也不好再多说些什么了。于是，德布罗意就这样“攻读”下了他的博士学位。而按照当时欧洲的学术传统，朗之万则将德布罗意的博士论文印成若干份寄到了欧洲各大学的物理系。

大约所有人都以为事情会就此了结，多少年以后德布罗意那篇“很新很有趣”的博士论文一定也会被埋藏到了档案堆里无人问津。

但是，在朗之万寄出的博士论文中，有一份来到了苏黎士大学。

这是1926年年初的苏黎士，春寒料峭。苏黎士位于阿尔卑斯山脉北部，苏黎士湖西北端，利马特河同苏黎士湖的河口。市区被利马特河分

为东、西两岸，也分成新城和旧城，它们之间有迷宫般的羊肠小道连接。旧城区在河北岸，分为上村和下村，布满了大大小小的精品时装店，酒吧，咖啡屋，古玩厅等。而苏黎士人口只有30万，因此是一个精美绝伦的小城。

当时在瑞士联邦工业大学主持数学活动的是外尔（Hermann Weyl, 1885—1955），主持物理学术活动的教授是德拜（Debye, Peter Joseph Wilhelm），德拜收到朗之万邮寄过来的这份博士论文后，将它交给隔壁大学一位已经年届中年的教师。

这位教师接到的任务是在两周后的学术例会上将该博士论文报告一下。

6 薛定谔：遗情书

(1)

学长的墓碑矗立在眼前，坟上有寂寞的小花慵懒地开放，这是一个寒冷的清晨，中央公墓里根本没有什么人，空气也仿佛被凝固，几滴露水从高耸的树上滴落下来，发出摔碎的声音，这个清晨格外寂静。

墓碑上的照片依然清晰，是一个大胡子的中年人忧郁的眼神。

碑文显得很简单，上面写着一个数学公式，或者说是一个物理公式，学弟有些看不清楚到底写的是什么。

学弟薛定谔（Erwin Rudolf Josef Alexander Schrödinger, 1887—1961）站在玻耳兹曼的墓前，心情有些压抑，因为他看到的那个博士论文也就是那个富二代德布罗意的博士论文中抛出了一个强烈模仿爱因斯坦的观点，认为质量不等于零的电子也同时具有波动性和粒子性，因此德拜叫薛定谔写一个方程出来描述电子的行为，但这事情根本就不是人干的，除非自己是疯子才能干出来吧，薛定谔暗忖。所以他连夜从苏黎士跑到维也纳来，跑到中央公墓来看看，自己都40多岁的人了，做物理如此不成功，不由得悲从中来。



绘画：张京

学长玻耳兹曼是自己的老乡，也是自己年轻时候做物理的偶像，现在学长已经死了十来年了，自己跑到这里来，也算是祭奠一下自己年少时的梦想。说起来，薛定谔一直把玻耳兹曼当作偶像。他也一直梦想着自己能成为一个偶像，但现在看来好像毫无迹象，而四十不惑，薛定谔觉得自己在物理上根本没有表现出什么天才，自己的一辈子，恐怕也就是这样庸庸碌碌地过去了。

薛定谔看清楚了墓碑上的字，这是后来才刻上去的，写着 $S=k\ln W$ 。薛定谔知道，这个公式表示在这个世界上，孤立体系熵 S 是增加的（也就是说，孤立体系总会趋向于越来越混乱）。这被称为最大熵原理，是这个宇宙运作的基本规律之一。

可是，在这个世界上，还有一套力学规律，同样非常简单，那就是最小作用量原理，所有的轨道当中，粒子选择作用量最小的轨道……

事情也许很简单，要成为偶像，也许必须模仿学长的剑法。

薛定谔想到这里，灵魂开始震颤。对了，就这样干，他的唇边露出微笑，喃喃地说：“让作用量也等于一个对数！”

在电光火石之间，薛定谔的脑子里突然浮现出一个美女雪白的大

腿，他有些惊慌失措，连忙蹲下来，捡起坟头的一个松枝，在泥地上写下类似的公式。

$$S = -i\hbar \ln \psi$$

其中S是作用量， ψ 是一个暂时不清楚的东西，地位大致等价于墓碑上的那个W。

天啊，这个山寨版本的公式和墓碑上的那行字惊人地相似。薛定谔仔细地看了一下墓碑，觉得自己这一招偷梁换柱做得简直天衣无缝，如果他自己不说这个对数，根本没有人知道这个东西是怎么来的，虽然这好像是抄袭了学长的什么东西，他的脸上浮现出一丝尴尬。幸亏没有别人看到，薛定谔站起来，用脚把写在地上的文字匆匆抹去，然后他弯腰朝学长的墓碑鞠躬下去，像是在感谢这地下的死者给他的冥冥之中的保佑。

(2)

维也纳的三剑客，薛定谔是多情剑客。

薛定谔其实是一名诗人，他从维也纳回去的时候，身边已经多了一位少女——这不是他的妻子，而是他最喜欢的前女友，一个曾经的可爱小萝莉。他虽然表面上在做物理，但内心里有波涛汹涌的情感——他甚至还思考过生命是什么这样的终极问题。

这是1925年的圣诞，阿尔卑斯山上直插云霄的皑皑白雪，吸引了各地的旅游度假者。薛定谔他们来到滑雪场，忘却了自己已经是一个已婚男人，这次他约了前女友来这里滑雪，顺便把德拜让他写的波动方程也写一写。自从分手以后，他对她是万分想念。前女友来了，很多年没有见，她还是那么美……事情就是这样简单，著者无法用前女友的笔调写下《遗情书》内的种种细节，总之，薛定谔在这个圣诞节就好像中了黯然销魂掌，如果说来这里滑雪会摔断腿，薛定谔也不会不来。

和她在这里，真是爽。他和她疯狂地接吻，疯狂地爱抚.....简直是惊天地泣鬼神。薛定谔趴在她身上，咀嚼着德布罗意的思想，德布罗意的思想其实不算是一个创新，原因在前面已经写过，因为早在1905年，也就是20年前，爱因斯坦就已经说过无质量的光子具有波粒二象性。而现在20年后，这个德布罗意说，那么.....也许.....有质量的电子也有波粒二象性，好吧，现在薛定谔觉得，既然电子是粒子，那么它肯定有一个轨道的作用量 S ，这个作用量和波动函数之间的桥梁是什么呢？他猜想，这个桥梁是一个对数，至于为什么是对数，薛定谔看着前女友那张迷人的沉醉笑靥，也想不清楚——也许是那个死去学长的启迪吧，反正就按照对数来做吧。

总之，薛定谔决定把它用到原子体系的电子的描述中去。

他知道作用量 S 是要满足哈密顿-雅可比方程，这是一个偏微分方程，已经被使用了百年，早已经不新鲜，现在新鲜的东西，就在那个对数里。

(3)

薛定谔的心里已经有谱，就是要把粒子性和波动性结合起来。

$S = -i\hbar \ln \psi$ 是他的秘密，他仿佛心里有鬼似的，趁着前女友睡觉的时候，拿出纸张，在上面写起来。粒子性和波动性已经结合起来了，就是他的这个怪招 $S = -i\hbar \ln \psi$ 。问题在于，如何做才能说服德拜他们呢，德拜是他在大学里的同事，也是物理学教授。薛定谔知道得很清楚，德拜不是那么好糊弄的。

那只好把这个 $S = -i\hbar \ln \psi$ 代入经典粒子运动的哈密顿-雅可比方程了。薛定谔知道，这样做在逻辑上还是可以的，就是说法非常像民间科学家，他的脑子有些混乱，因为波粒二象性对于电子来说，真是确有其事吗？他甚至有时候觉得德布罗意是一个混混，这种电子的波动性真的

存在吗？作为一个诗人，薛定谔其实是一个充满了怀疑的颓废，他对别人是非常不信任的。

没有办法了，死活就这一招，把那个对数代进哈密顿-雅可比方程了以后，床上的女朋友翻身，嘴巴里似乎在嘟囔什么，应该是在梦呓吧。薛定谔审视了一下眼前横陈的玉体，觉得一片明媚，见她还没有苏醒，格外紧张，继续写下去，利用变分法，最后他得出了一个波动方程，最终形式是这样的：

$$-i\hbar d\psi/dt = \Delta\psi + (E - V)\psi$$

这时候天色已经渐渐暗下来，皎洁的月光照在雪地里显得格外柔和。薛定谔看着写在纸张上的那个波函数方程，觉得有点可笑，他不知道自己是不是真的在做物理。虽然此方程的样子很可爱，但左右也不完全对称，好像是一个热传导方程的样子，而不是完全的波动方程。但他知道，要判断这个方程成立不成立，只要解答一个现实问题，如果能从这个方程中得到原子的光谱和能级，那么此方程很可能是对的。因此，当下最重要的事情，是求解这个方程，但此方程奇形怪状，涉及以前讲到的“斯图姆-刘维尔方程”这个系统问题，真的不好解，薛定谔身边没有参考书，于是又着急起来。转眼又看了一下旁边睡着的美女，再次觉得这个世界真的很纯真。

纯真在月光下裸奔。

为了求解这个日后名震江湖的薛定谔波函数方程，薛定谔心想，既然是自己造出来的方程，自己就要负责，现在方程中的那个三角 Δ 叫做“拉普拉斯算符”，代表了某种微分运算。 E 是体系总能量， V 是势能，在原子也就是 $-e^2/r$ 。

而原子的能量 E 是离散的，那么，怎么样才可以得到离散的能量呢？拉普拉斯算符的本征值问题，薛定谔多吃了几年饭，也是知道的，具有离散的本征数值，那么，也许，可以把 E 和拉普拉斯算符组成的那

部分，解出来。可是，方程的左边，还有与时间相关的项呢？怎么处理？薛定谔心想，这个波动函数，实际上应该是与时间无关的稳定状态，所以，时间部分一定可以分离出来，于是，他就丢弃时间部分暂时不管。

这个时候，前女朋友翻身，亲昵地叫了一声，见薛定谔若有所思，就一把从背后抱住他，薛定谔感觉自己背后被两团软软的东西压住，就放下了手上的笔.....

(4)

回到大学里，薛定谔找到希尔伯特他们写的数学物理方法的参考书，开始把自己写的那个方程完整地解答出来，发现整理出来的结果真的很有趣，居然能够给出氢原子的能级，而且与光谱的实验观测数据全部对上了，这真是太神奇了。他于是报告给德拜，后者也用薛定谔方程完整地解了一下氢原子中的电子的能量，发现果然能量是分离的离散数值，惊讶得下巴都要掉下来了。

薛定谔则一鼓作气，写了四篇雄文，奠定了波动力学的基础。

薛定谔的方程一出世，几乎全世界的物理学家都为之震动，感到物理学真的不一样了。爱因斯坦说：“.....您的想法源自于真正的天才。”实际上，爱因斯坦对薛定谔的评价一点也不夸张。很多年以后，大学物理学教授们还是不知道薛定谔到底是怎么样推出他的方程来的，只有薛定谔自己知道，这其实是前女友的《遗情书》，但这书中，还有自己年少时候偶像，死去学长的影子。

外一篇 晶体光栅和电子的波动性

1926年，一位美国的实验物理学家登上了去英伦的船。他叫戴维孙，有一个实验助手，叫革末，他们喜欢用高速电子流去轰击各种金属

样品，然后再把散射电子的能量数据记录下来。这几年来，他们不断地干这样的事情，可以说已经干到山穷水尽的地步。

他们发现从金属靶上发射的“二次电子”有少数具有与一次电子相同的能量，显然是在金属反射时发生了弹性碰撞。——他们只关心这些经过弹性碰撞出来的电子。

可是问题在于，有时候这些弹性碰撞出来的能量不变的“二次电子”的角度分布有两个极大值。他们试图仿照卢瑟福 α 散射实验用原子核对电子的静电作用力解释这一曲线——卢瑟福散射中，被原子核散射出来的“二次电子”的角度分布曲线并没有两个极大值。

戴维孙这个时候其实已经把两个事情混淆起来。

1. 电子在卢瑟福的核电场弹性散射
2. 电子在金属晶体上的衍射

不过，因为没有人告诉戴维孙电子是一种波（波才会衍射，具有在不同角度上的衍射峰，单缝衍射的光强分布是一个sinc函数的平方，sinc函数的定义是 $\sin x/x$ ——这个函数非常重要，因为他是矩形门函数的傅里叶变换，所以经常出现在电路和光学的各个角落），所以美国物理界相安无事，只有戴维孙和革末两个人为这些实验现象苦恼。因为物理实验有时是有非常大的误差的，保不准在什么地方会出错，所以，有些物理现象并没有真正的物理含义，而仅仅是仪器或者样品出现问题而引起的。

这次去英国开会，不知道能听到些什么新东西。戴维孙站在船头，脑子里一片混乱，去年的实验现象更加费解。到了英国，这些科学家们在牛津大学开会。会议由著名的德国物理学家波恩（Max Born, 1882—1970）主持，他提到了德布罗意波。德布罗意波？戴维孙以前闻所未闻，他立即联想到了自己最近获得的两处尖锐的峰值的实验数据……真是一语惊醒梦中人，这很可能就是德布罗意所预言过的电子衍射！

这个晚上，戴维孙在英国睡不着觉了。脑子里反复在想：“电子？

是波？ $\lambda=h/p$ ？晶体相当于光栅？衍射了？”

光栅是具有周期性结构的镜子，一般来说，在一个毫米的距离上刻有几百到上千条凹槽。这最初是弗朗禾费发明的，用来对可见光做分光实验——就是把白光分成七色光，但光栅不同于棱镜，光栅还能把同一波长的光在不同角度分配能量。

因为电子的波长比可见光要短很多，跟X射线一样，光栅分光能力对它已经不起作用。大自然鬼斧神工，自然界里还有其他周期性的结构，比如晶体就是很好的“光栅”。1866年布拉维得到了14种晶体的点阵分类，后来已经由狄拉克的大舅子威格那开始把群论的思想引进到晶体这种具有高度对称性的东西里来了。劳厄对X射线的衍射做了很深入的研究，比较简单的关于晶体衍射方程则是布拉格方程。

既然话已经说到这里，我们不妨继续多说几句。

光栅的理论分辨率是与每毫米的光栅凹槽的数目成正比的。但对于X射线和电子来说，这个分辨率还是不够的，因为凡是光栅，都是人做出来的，对电子波来说，凹槽与凹槽之间的距离还是太大了，电子根本就表现不出波动性来。晶体可以用来充当光栅的角色。目前在中国就有几个分析仪器公司能够生产X射线衍射仪，用来做物相分析。

废话少说，戴维孙回到美国，准备了很纯净的单晶镍，在1927年和革末一起出色地再次完成电子波动衍射的实验。

薛定谔发现的方程就是用来描述这种波的。

7 十年前的玻尔

(1)

薛定谔是在39岁这一年，写出了薛定谔方程的，他写完以后，想到了很多人，比如爱因斯坦，德布罗意，也想到了另外一个人，那就是玻尔（Niels Henrik David Bohr，1885—1962）。

1911年，玻尔26岁，这年5月，他以长篇论文《金属电子论的研究》获哥本哈根大学哲学博士学位，旋即得到卡尔斯伯基金会资助出国一年；他选择了英国剑桥，准备在J.J.汤姆孙（Joseph John Thomson，1857—1940）指导下继续研究金属理论——汤姆孙这个时候正在做世界上第一台质谱仪器，对金属兴趣不大。

卡文迪许实验室的头头，电子的发现者，诺贝尔奖得主J.J.汤姆孙一开始十分热情地接待了玻尔，两人还是促膝长谈，大有相见恨晚的感觉。J.J.汤姆孙收下了玻尔带给他的一篇论文，并把它放在自己的办公桌上。

和大部分职场的面试差不多，一切看来似乎十分顺利。但事实上，玻尔的论文一直被闲置在J.J.汤姆孙的桌子上，他根本没有看过一个字，主要原因是因为当时刚从学校毕业的玻尔没有社会经验，当面指出了J.J.汤姆孙的著作《气体中的导电》里的一些错误——俗话说文如其人，这样做确实有点打脸的感觉，于是实际上他惹恼了高傲的英国绅士。不管怎样，剑桥对于玻尔来说，实在不算一个开心的地方，他明显受到冷遇。除了在一个足球队里大显身手之外，这所举世闻名的大学似乎让玻尔觉得没有什么是一值得一提的。

这个时候英国有另外一位物理学家，新西兰人卢瑟福（Ernest Rutherford，1871—1937）已在曼彻斯特通过 α 粒子散射实验确证了原子

核的存在。事情是这样的，早在1897年，J.J.汤姆孙在研究阴极射线的时候，就发现了原子中电子的存在。但是，原子的空间结构究竟是怎样的呢？那时完全缺乏实验证据，J.J.汤姆孙于是展开自己的想象，勾勒出这样的图景：原子呈球状，带正电荷。而带负电荷的电子则一粒粒地“镶嵌”在这个圆球上。这样的一幅画面，也就是史称的“葡萄干布丁”模型，电子就像布丁上的葡萄干一样。但是，1910年，卢瑟福和学生们进行了一次名留青史的实验。他们用 α 粒子（带正电的氦核）来轰击一张极薄的金箔，想通过散射来确认那个“葡萄干布丁”的大小和性质。但是，极为不可思议的情况出现了：有少数 α 粒子的散射角度很大，以至超过 90° 。对于这个情况，卢瑟福自己描述得非常形象：“这就像你用十五英寸的炮弹向一张纸轰击，结果这炮弹却被反弹了回来，反而击中了你自己一样。”——这个著名的实验标志着人类开始进入原子核能探索的时代。

1912年3月，玻尔离开剑桥赴曼彻斯特，跟卢瑟福学习原子结构理论，主要是卢瑟福的所谓原子的太阳系模型——原子之内是一个非常宏大的世界，电子像行星绕太阳公转一样绕着原子核作圆周运动，这个经典图像虽然很简洁，但有一些难以克服的理论问题，比如电子作圆周运动会发出的同步辐射问题。同年7月，玻尔撰写了一论文提纲交给卢瑟福，后人称之为《曼彻斯特备忘录》——在这里他开始有了一个强行的规定，规定原子内的电子在作圆周运动中是不会发生同步辐射的。当时原子光谱线的规律早被找到了。事情分成两个部分。



绘画：张京

1. 巴尔末发现：氢原子的光谱线的波长的倒数正好是与两个自然数倒数的平方差成正比。

2. 莫塞莱发现：X射线光谱线的特征波长的倒数与原子序数的平方成正比。

巴尔末是一个在瑞士的中学数学老师，其实是半个民间科学家，但他的发现需要很强的洞察力，能够从复杂的光谱数据中找到规律，这种工作其实一般人是绝对做不出来的。这需要盯着一堆貌似杂乱无章的数据看很久很久。

而莫塞莱发现的规律虽然是针对X射线，但X射线其实也是光谱线，这也说明了很重要的一个线性关系。他搞出这个线性关系以后，就可以修正元素周期表里错排的项。

总之，这两个人的发现是非常独立的两个侧面。弗朗禾费时代以来，人们已经可以很完善地记录谱线的波长，但这些光谱波长之间的排列到底有什么规律，没有人晓得。这些经验规律背后的物理到底是什么呢？玻尔就是研究这些问题的。

1912年7月底玻尔离英回国，8月1日，玻尔与玛格丽特结婚，9月1

日，开始在丹麦哥本哈根大学任教，1913年2月，玻尔的注意力“突然转向”原子光谱的规律，从而大大发展了原有的关于原子结构的看法，这年的7月、9月和11月，他以《论原子构造和分子构造》为题，在英国的《哲学杂志》上分三次发表长篇论文，奠定了他的原子结构理论的基础。1914年，29岁的玻尔应卢瑟福之聘，到曼彻斯特任讲师。所以那几年玻尔在哥本哈根和曼彻斯特两地来回穿梭，已经搞不清楚自己到底是要留在英国发展，还是回丹麦发展。

(2)

1916年的一个清晨，从英国到丹麦，是一片汪洋，海水暗蓝，乌云已经开始密布，玻尔坐在甲板边的栏杆上，邮轮下巨大的螺旋桨激起飞溅的浪花。海浪渐渐地大了起来，海天在远处连成一色，故国就在远方，玻尔心里已经万分焦急，仿佛是要去会晤情人。突然，海面上一个巨浪掀来，船体剧烈地震荡起来，玻尔差点倒栽进海里喂鱼，吓出一身冷汗，他忙从栏杆上下来，看来，海上似乎又有一场大暴雨了。玻尔这是从曼彻斯特回来，这一次，他要脱离卢瑟福，自立门户。若干天后，玻尔上岸了。玻尔说了这样一句话：“我们现在回来了，丹麦将大不同。”同样在这一年，在德国的爱因斯坦已经学明白了黎曼几何学，他把时间和空间组成一个弯曲的流形，这使得他们两人在日后成为物理学江湖上的东邪西毒奠定了基础。一年以后，玻尔建立了哥本哈根研究所。在这一年，奥匈帝国的薛定谔离开维也纳大学前往军队成为一名炮兵军官，他服役于一个偏僻的炮兵要塞。薛定谔这个时候还没有结婚，这年他已经30岁。

玻尔只不过比薛定谔大两岁，薛定谔要到39岁才大器晚成，玻尔早在1913年28岁的时候已经在物理学界有了一定的名气。1913年他得到了一个“环路积分”，这个环路积分被称为量子化条件，也就是轨道角动量

的量子化条件，这个积分其实也是玻尔为了说明原子的分离能级，硬生生的一个假设，但没有想到这个很僵硬的假设也在物理学上被证明是对的，至于为什么会这样，玻尔是不清楚——这就好像在中国一般的高新企业在创业初期，凡是有点带有高科技性的产品，无论这个产品的设计与界面做得多么烂，总能卖出去一两台，这样的企业往往能攫取第一桶金然后越滚越大，玻尔也是在差不多的情况下攫取了人生的第一桶金。

玻尔壮怀激烈地回到了自己的祖国——丹麦，开始筹建一个研究所。1917年这个研究所就矗立起来了，这就是哥本哈根学派的大本营。玻尔心气很高，他对灯发誓要干一番宏大的事业，他对别人说：科学没有国界，但科学家是有祖国的。

(3)

电子的运动到底有没有轨道呢？

玻尔做博士后期间的老板卢瑟福认为，电子运动是有轨道的，并且轨道是圆的，而且轨道半径是非常任意的。电子在原子内运动就像地球在太阳系内运动一样。但很明显，同步辐射会让这个小太阳系模型不稳定，这就好像一个人如果拉着一件刚从水里捞出来的衣服转圈，衣服中的水就会被甩出去一样。经典电磁理论预言，这样的体系将会无可避免地释放出辐射能量，并最终导致体系的崩溃。换句话说，卢瑟福的原子模型是不可能稳定存在超过1秒钟的。

这是一个巨大的问题呀。玻尔也陷入了沉思，有一天，他终于明白了一个道理，那就是原子内电子的轨道必须和光辐射的能量一起来考虑。因为在这之前，德国的普朗克已经得到一个重要的内容，就是光辐射振子的能量是离散的。这一点也给了玻尔一些启发。因此如果把辐射振子的能量看成了经典相空间（平面）上的轨道，那么很容易推出来，只有在相平面上特定半径的一些轨道才给出辐射振子的离散的能量。于

是，玻尔就得到了前面说的“环路积分”。把这个量子化推广到原子内的电子轨道，也是同样道理。

通过同样手法的简单计算，就可以知道，卢瑟福所说的圆轨道，轨道的半径并不是任意的，而只能是一些特定的离散数值。也就是说，给你一个原子，它内部只有特定半径的轨道可以让电子去奔跑。而不同半径的轨道能量是不一样的，轨道之间的能量差正好就是光辐射的能量。这些电子的轨道就好像北京城的二环、三环、四环、五环和六环一样，车子只可以在这些环路上奔跑，北京是不存在4.5环这样的高速路的。在不同的环线之间，汽车可以飞过来，比如直接从二环的积水潭桥飞到四环的保福寺桥。也可以直接从三环的北太平庄桥飞到二环的官园桥。总之，汽车只能走环线，而不可以走新街口外大街这样的有红绿灯的路，如果要在环线之间切换，就必须让汽车飞起来。

这就是玻尔的原子模型。

但问题是，为什么电子必须只能走这些离散的轨道呢？

玻尔没有办法解释，——因为实际上电子是没有轨道的！！

(4)

那么，没有轨道，电子是怎么运动的呢？这就是1925年薛定谔写出的那个波动方程描述的。电子其实是一种波动。十年前的1913年，玻尔当时还没有超越时代，他觉得，电子是有轨道的，但只能取一些离散的轨道。不同的圆形轨道可以用自然数1, 2, 3, ..., n 来标记。玻尔这个时候还算不上一个大物理学家，他强行规定，电子只能在特定的轨道上运动。他的这个做法其实是非常野蛮的，但不可否认，他能够用同一种语言把光谱和电子轨道联系起来是一个很大的进步。真正完美的计算需要再等13年，计算是出自情圣薛定谔，这个已经在上一章讲过了。

薛定谔写出波动方程以后，元气大伤，在床上躺着的时候他总是想

一个问题，这个波函数究竟是什么意思呢？虽然方程已经写出来，也能够算出氢原子内电子的分立的能级。但这只不过是事情的一个侧面，能级分立其实出自微分方程本身的结构，而作为微分方程里的主要未知量，波函数包含什么样的物理，却是很费解的。

上次在滑雪场确实是春心荡漾，可惜现在回头想起来，难免有些空虚，更加重要的事情还没有干呢，因为波函数的方程虽然写出来了，但这到底意味着什么呢？电子的运动没有轨道，这跟波函数有什么关系？ $S = -i\hbar \ln \psi$ ，在原子中，电子的每一个可能的轨道，都有一个作用量，那么，电子到底是怎么运动的呢？波动？在哪里波动？

波心荡，冷月无声，窗外一片寂静。

外一篇 函数方程：对应原理

(1)

前面已经讲到，1925年薛定谔关于波函数的文章充满了鬼打架的风格，那时候玻尔已经组建了一个哥本哈根学派，那里的很多人包括玻尔在内，对这种风格莫衷一是。薛定谔的方程里，没有物理上可观测的量，玻尔则开始处处强调物理上可观测的内容，比如光的频率，光的强度，在薛定谔的方程里，却没有这些。但薛定谔的方程可以解出氢原子的能级，这已经足够了——薛定谔的计算结果，和玻尔八年前的结果殊途同归。

八年前的结果可以被重现，自然说明薛定谔已经走到了一个绝妙的境地。

画鬼容易画人难，做学问永远是这样的。

量子力学在发展之初，也有这种画鬼思潮的痕迹。这种画鬼的思考方式起源于德布罗意，在薛定谔成为绝响。

我们暂时告别薛定谔老师，不再探讨他武功的路数，而转而去看，一个真正的物理学家，是如何做物理的。

(2)

玻尔年轻的时候，解决了氢原子的能级问题。他的思路是非常自然的，不会让任何人觉得吃惊。这个思路的核心就是所谓“对应原理”，这个原理成为海森堡后来最厉害的思想武器。实际上，对后来者来说，对应原理是一种真正的物理方法，换句话说——这是物理学家做事情的一般方法，先假设，再求证。

在玻尔的原子模型里，电子在不同的轨道上运动，这些轨道可以用自然数 n 来标记。读者们一定要注意了，其实轨道是不存在的，但物理学家不可能先验地知道轨道不存在，所以，玻尔的思路是非常完整的。在经典力学里就可以知道，不同轨道的能量不一样，可以把第 n 个轨道的能量记为 $E(n)$ 。

因为 n 是一个整数，所以 $E(n)$ 是一个未知的数论函数。

玻尔认为，电子可以在不同的轨道之间相互跳跃。这被称为跃迁——类似于股票市场中的那种跳跃，比如，打开任何一个股票交易软件，今天的上证指数到了收盘的时候已经有了一条轨道，假设收盘在2890点，那么，明天早上开盘不一定是在2890点，有可能跳空高开，比如在2920点开盘。

电子的轨道也是如此，从能量高的轨道跳到能量低的轨道，电子的能量肯定要释放出来，这就满足如下的能量守恒方程。

$$E(n+m) - E(n) = h\nu(m,n)$$

这是一个函数方程， $\nu(m,n)$ 表示光谱线的频率，这实际上是一个类似于 $F(n) + F(n+1) = F(n+2)$ 这样的被称为菲波那切数列的函数方程。菲波那切数列的函数方程的目标是求出 $F(n)$ 的表达式。同

样道理，玻尔要求出 $E(n)$ 的表达式——这个表达式与整数 n 有关系，具有能量量纲。

(3)

$$E(n + m) - E(n) = h\nu(m, n)$$

这个方程的左边是2个能级之间的能量差，而右边是放出光子的能量。这个方程可以解释世界上所有的线光谱，所以，求解它显得尤为重要。

这个方程的右边是可以观测的，就是光的频率（波长可以通过用正弦机构带动旋转的光栅组成的单色器测定，频率是波长的倒数）。但左边是不能观测的原子的能级。求解的关键自然在于确定右边的函数形式。

这个时候， $\nu(m, n)$ 的表达式是不能通过眼睛看出来的，必须要有一个假设来支撑它。玻尔使用了如下的假设，被称为对应原理：当 n 很大同时 m 很小的时候， $\nu(m, n)$ 作为放出光子的频率等于电子在圆周轨道上运动的圆周运动频率的 m 倍。

高中学生都知道，一个电子作圆周运动的时候，它的角频率是圆周运动的速度和半径之比。为了计算方便，可以取 $m=1$ ，那么我们可以得到

$$E(n + 1) - E(n) = h\nu(1, n)$$

对应原理说：

$$\lim_{n \rightarrow \infty} [E(n + 1) - E(n)] = h\nu$$

其中 ν 是经典圆轨道的频率，这个频率是和能量 E 的 $3/2$ 次方成正比的（高中物理）。

所以，我们有如下表达式：

$$\lim_{n \rightarrow \infty} (E_{n+1} - E_n) = CE_n^{\frac{3}{2}}$$

其中C是比例系数，是常数。

也就是说E（n）对n的导数正比于E（n）的3/2次方，可以推出，E（n）正比于n⁻²。这样就解出了氢原子的能级表达式。

（4）

对应原理解出的氢原子的能级非常符合观测到的光谱数据，所以，这个原理成为思想的利器。玻尔在这个时候开始成为一个真正的物理学大师。真正的物理学大师不需要太多的数学，只需要在非常恰当的时候做出一些恰如其分的物理假设。在这个故事里，玻尔为了解出一个函数方程做了一个当n无穷大情景下的渐近假设，这个假设看起来也是非常合理的，因为他只不过要求一个量子系统在量子数很大的时候非常接近于经典系统。

对应原理把量子力学拉回到经典力学，这是必须的，因为量子力学在某种意义上是一门画鬼的学问，但最后必须要能回到人的世界。

8 物理思想集大成者：爱因斯坦

(1)

薛定谔写出波动方程以后，不但想起了玻尔在10年前的工作，他也想起了爱因斯坦。爱因斯坦是波动学说的支持者，薛定谔也确实是沿着爱因斯坦和德布罗意的路线走过来的。爱因斯坦在6年前，就已经是一名处于媒体聚光灯下的物理学大师，他被认为是修正了牛顿绝对时空观的人。

而这一切还要从1919年说起。

那是在1919年5月29日一个宁静的清晨，太阳刚刚升起。巴西南部一个叫做索不拉尔的偏僻村庄，偶然有一两声狗吠，草原和麦田在村子外蔓延，一直远到天边。

但是，与往日的宁静不同，这天村子门口来了一队行装古怪的外国人。有的操着蹩脚的葡萄牙语向村民们问好，有的打着手势借水喝，更多人在忙着架设天文望远镜和照相机。

看得出来，这是一群来自欧洲的天文学家，他们似乎要来这里拍摄有关日食的照片——日食，又作日蚀，是一种天文现象，只在月球运行至太阳与地球之间时发生。此时，对地球上的部分地区来说，月球位于太阳前方，因此来自太阳的部分或全部光线被挡住，看起来好像是太阳的一部分或全部消失了。日食分为三种，包括日全食、日环食与日偏食，其中较罕见的是日环食。

中午时分，灿烂的阳光渐渐褪色，一个黑影渐渐地遮蔽了太阳。围观的人群开始骚动了起来，经年不遇的日全食终于又发生了。

天空变得越来越暗淡，风也大了起来，吹得一个教授模样的人面前的天文望远镜的镜头微微颤抖。

领队的教授吩咐手下的一个摄影师在这微熏的黑暗中按下了照相机的快门。他的目光明显流露出期待，看他的样子，似乎是在给天空外的遥远的星星拍照。

满脸写着狐疑的村民们在边上个个敛气屏声，一言不发，一会儿看看实验人员，一会儿又望望黑暗的天空。很快，照相完成了，教授模样的人转身拿着照相机走进了临时搭建的大帐篷里，天空渐渐露出了一点鱼肚白，看来日全食也正在慢慢过去。大帐篷的一个角度被一块帆布隔开，那里很像是一个洗照片的暗室，教授模样的人进去以后，将湿漉漉的底片放在灯下时，很快教授先生的手连同大胡子都激动地抖了起来：空间是弯曲的？光线被弯曲了！牛顿那个僵化的平坦的没有弹性的空间，被粉碎了！他深吸一口气，睁大了眼睛，再仔细地看了一下照相的底片。

在两张重叠的底片上可以清晰地看到一条笔直的星光在穿过阴影中的太阳时，竟然发生了偏转，偏转角可以被测量出来，是1.7秒……

(2)

英国皇家学会的大厅，座无虚席，第一次世界大战的硝烟味似乎还在空气中弥漫，但科学家们还是闻到了一股新的气味，好像与往常不同。



绘画：张京

过了一会，大会主席汤姆孙爵士走上讲台，他是发现了电子以后得到诺贝尔物理学奖的，是一个对微观世界非常了解的人，但他对宏观世界和万有引力定理，却不太了解，因此，当读到大会的主要议题的时候，显得有半点的怀疑，慢慢地，他扶了扶眼镜：“现在，我宣布，今天这次大会的议题是‘广义相对论在天文学上的验证’！”

底下的学者们听到这里，开始骚动不安，纷纷交头接耳起来，底下的声音也越来越大，有的甚至不等主持人的召唤隔着橡木桌就站起来大声辩论……

“广义相对论？什么东西？爱因斯坦是那个德国人吗？”一个说。“相对论是错的！什么叫广义相对论？还可以验证？”另一个附和道。

汤姆孙摇了摇手中的铃铛，示意大家安静，“各位，下面我们请天文学家爱丁顿勋爵谈谈他的看法。”语音未落，一个风度翩翩的绅士已经站起来，出现在众人的视野里。

这就是著名的天文学家，爱丁顿勋爵，这次天文测量的总领队，他当时去的是另外一个观测点，一个在非洲的小岛，他领导的观测小组在

那里做了同样的日全食观测，数据也差不多符合爱因斯坦广义相对论的预言——所谓符合就是误差在1倍左右，因为靠照相拍摄星星的视位置的方法其实验的精密程度实在让人不能恭维——这就好像隔了1光年的距离让一个狙击手去瞄准一颗行星，谁能说清楚到底瞄准了没有呢。

爱丁顿勋爵上台以后，下面顿时安静了下来——因为大家都认识他，知道他是一个令人敬畏的天文学家，甚至是地球上唯一一个自称能懂爱因斯坦广义相对论的科学家。

爱丁顿说：“我们的实验，是在日全食发生的时候，在地球上的两个相隔遥远的地方，同时做的……铁一般的事实……光线确实被太阳的引力场弯曲了……实验的结果与爱因斯坦博士的计算结果完全一致……”

他这个讲话虽然有不少夸张的成分，尤其是“完全一致”的这个论断非常像中国的专家，属于典型的夸夸其谈，但因为别人都不懂这究竟是怎么回事，于是也就只好再说什么。

媒体也马上获得了这一项重要的实验进展，第二天的《泰晤士报》头版头条的报道是《英国天文学家支持德国物理学家的理论：光线弯曲了，牛顿神话破灭》，报道说：“英国科学家支持了德国科学家的理论。爱丁顿爵士宣称，目前地球上只有3个人懂得爱因斯坦博士的广义相对论，而且，爱丁顿爵士还想不到第3个人到底是谁……”很快，从伦敦街头的商贩，到曼彻斯特地下巷道里面目黝黑的煤炭工人，都隐约知道了科学界最近发生了惊天动地的事——牛顿理论，不行了！现在出来了一个德国人，推翻了牛顿理论。在大西洋彼岸的纽约，《纽约时报》的头条标题是“俄国爆发革命”，但接下去以更大的标题写道：“爱因斯坦的胜利”，报道说“恒星在不在它们应在的位置上出现，似乎不必担心。但说明牛顿万有引力定理是错误的，而应该被爱因斯坦的广义相对论所取代。牛顿理论的崩溃正在引起公众们对科学真理的恐慌情绪，很多人已经开始怀疑九九乘法表的正确性，学生们则开始拒绝作几何题，

又据不愿透露姓名的消息人士透露，爱因斯坦在把他的著作交付出版商时，警告说全世界仅有12个人懂相对论，但出版商乐于承担这个风险.....”在浪漫之都巴黎，小资们的沙龙里“相对论”马上成了最时髦的词语。雍容华贵的妇人们可以一边抚摩着怀中的哈巴狗，一边和女友们眉飞色舞地谈论最新的“相对论”——时间弯曲，空间弯曲，星光弯曲，一切都在扭曲，这一切如同在谈论昨夜刚上演的歌剧，栩栩如生。在柏林，官方正在为难是否宣传这位并不是日耳曼人而是犹太人的传奇科学家的时候，大街小巷的啤酒馆里的人都在神秘而兴奋地谈论着爱因斯坦和他的相对论。是的，自从一战以德国的惨败而告终后，很久没有这样激动人心的话题了。一个德国人能受到战胜国英国的推崇，真是少见。一夜之间，小学生们也把爱因斯坦那著名的象征着相对论的公式 $E=mc^2$ 写入了家庭作业的练习簿。第一次世界大战的硝烟刚刚散尽。为民族主义所鼓动的人们在大喜大悲过后，却发现轰轰烈烈的一战除了大口径重炮，齐柏林飞艇，满目疮痍的建筑物和以百万计亲人充当炮灰以外，委实没有剩下什么。各地开始出现各种宣传相对论和爱因斯坦的漫画，有些漫画甚至把光线弯曲画成一个小偷在黑暗中作案，但因为光线弯曲，他还是被人发现了！



绘画：张京

对媒体与大多数人来说，爱因斯坦是一个耀眼的科学明星。他的思想博大深邃，而这一切又似乎是在一夜之间发生的。而真实的情景却不是这样的，其实早在1911年，爱因斯坦在《物理年鉴》上发表了题为《关于引力对光线传播的影响》的论文，论文中预言：当恒星的光非常接近太阳时，会因为太阳的引力而产生小小的偏离，而这种光线的弯曲是可以测量的。但这个时候，爱因斯坦还没有完整地考虑时空弯曲对光线的影响，他只考虑一部分有均匀引力场产生的光线弯曲。也就是说，引力场的存在使得我们人类就好像水中的鱼在看岸边的树，树会显得比真实的高度要高一些——这一切都是典型的光的折射效应。

(3)

对于当时的普罗大众来说，相对论到底是什么，是不可能搞明白的深奥理论。而对爱因斯坦来说，他从狭义相对论走到广义相对论，也花了10年的时间，这10年来筚路蓝缕，其中的艰辛与磨难，也只有他自己知道。

更早的时候，1907年，有人请爱因斯坦写一个介绍狭义相对论的综述文章，写这样的文章，使得爱因斯坦重新全面地审视了一下自己的理论和周围的世界。狭义相对论是在1905年建立的。当时的爱因斯坦依然在伯尔尼瑞士专利局，他经常坐在书桌边一边做一些专利审查工作，一边则想想自己的相对论理论，在他的狭义相对论中，他发现世界上的每一个人的手表，走动的快慢是不一样的，而且有的时候根本无法相互对准。正当他为关于时间的私有性感到非常神秘的时候，他又突然遇见了他一生中最快乐的思想——等效原理，“我正坐在伯尔尼专利局的桌旁，突然出现了一个想法，‘如果一个人自由下落，他将感受不到自己的重量。’”换一句话说，引力质量等于惯性质量。爱因斯坦把这个称为

等效原理。

这个原理其实很简单，就是在高中阶段的物理课本中，重力场 $G=mg$ 是可以被一个非惯性参考系的均匀加速度所等效的。这本来也没有什么新颖性，但是爱因斯坦比较深邃，他马上想到另外一个问题——引力场的能量到底是谁观测到的呢？比如一个人朝太阳掉下去，按照等效原理，在他看来，他没有感受到任何引力，相当于他没有测量到引力场的能量。这明显不同于电磁场的情况。因为电场与磁场是相互补充的，不同的人看到的电场和磁场的总能量是一个不变量——就算是在一个加速的系统里也是如此，无非就是有电磁波的产生，在这个时候总的能量还是可以计算出来的。因此，我们总可以说，如果一个静止的人看到一个静电场，在另外一个跑动的人看来，既有电场还有磁场，但他们两个人看到的电场和磁场的总能量是一样的，这就是说，电场和磁场的能量是可以局部到一点来谈论它的密度。现在问题变得不一样了，首先是爱因斯坦从来没有听说过引力场还有引力磁场这部分东西的存在，而且对应于电磁波的引力波也是没有听说过的，那么这个关于引力的理论到底应该怎么改造呢？引力能量有没有局部的密度？这在爱因斯坦心头是一个巨大的问题。

另外，按照这个等效原理可以知道，在自由下落的电梯里如果有一个手电发光，那么这个手电发出的光在电梯里的人看来走的是直线，但在电梯外的人看来走的却是曲线。也就是说，使得这个电梯自由下落的是万有引力，同样也能弯曲光线？虽然牛顿本人早在1717年出版的《光学》书中，也有一个附录，一共提了31个疑问。其中第一个疑问就是问引力会不会对光线有弯曲作用，当然那时候是无法解答这个疑问的，现在爱因斯坦同样遇见了这个问题。

为了计算光线到底被引力场弯曲了多少，爱因斯坦做了一些计算，因为这个结果必须与狭义相对论关于时间是私有的这点进行协调，爱因斯坦不得不把整个问题与时间和空间放在一起进行打包研究，在这个过

程中，他意识到，时间与空间作为一个整体其实并不是别的什么东西，它们直接就是引力场。

时空就是引力场！

这是爱因斯坦最震撼人心的思想。

这是广义相对论的基本内容，我们将在本书的第二部分集中介绍。

1915年，爱因斯坦发表了介绍他的广义相对论思想的论文。这篇论文的影响本可能因为战争而被限制在德国，但爱因斯坦的两位密友：洛伦兹（Hendrik Antoon Lorentz）和埃伦费斯特（Paul Ehrenfest）收到了论文，并且与荷兰天文学家德西特（Willem de Sitter）分享。



第一排从左至右是：爱丁顿和洛伦兹；第二排从左至右是：爱因斯坦、埃伦费斯特和德西特。

照片拍于1923年9月的荷兰莱顿天文台

德西特作为英国皇家天文学会的秘书随后将文章发给英国的天文学家爱丁顿，后者1906年到格林尼治天文台工作，1913年任剑桥大学天文学教授，1914年后兼任该校天文台台长。

31岁就当上剑桥大学天文台长的爱丁顿认为爱因斯坦的论文“美、优雅、有力”，“这篇文章让我数夜不眠”。

但在当时英国反德情绪严重，无法发表一篇德文的报告，爱丁顿就让德西特写了一系列文章来介绍爱因斯坦的理论。这样，在战争中最黑暗的时期，英国人也了解了爱因斯坦的思想。

因此，如果说爱因斯坦1911年版本的光线偏折理论并不完善，但到了1916年爱因斯坦的广义相对论的完整版本给出的光线弯曲的预言是广义相对论的几大预言之一，这个效应已经完全考虑了时空弯曲的影响。另外一个预言是水星近日点的进动。离太阳最近的行星是水星，那儿的万有引力场强最大，广义相对论的修正最明显，之前天文学家已经观测到水星近日点存在进动，也就是说，人们开始注意水星的公转轨道不是一个封闭的椭圆，但没有人可以解释这到底是为什么。既然轨道不是椭圆，有了广义相对论就知道，水星与太阳之间的万有引力势场不是严格地与距离成平方反比的库仑势。在经典的力学里，有一个所谓贝特朗（Bertrand）定理，这个定理说只有当中心势是库仑势或者谐振子势的时候，轨道才是封闭的。这个定理是重要的，因为它否认了其他势场里存在封闭轨道的可能性，哪怕是对库仑势的微小偏离。所以，当爱因斯坦的广义相对论对万有引力的库仑势做修正的时候，在理论上，椭圆轨道就一定不是封闭的了，近日点在空间中会渐渐移动。广义相对论虽然比较难以理解，但在这个椭圆封闭性问题上，结论也是很清楚了。因此牛顿的万有引力定律，那样美的一个定律，在引力比较强的时候，也是不对的。

这些就是20世纪20年代爱因斯坦已经成为大物理学家的思想内核所在，而他所做的物理工作其实还有很多是集中在量子力学方面，在那方面他的思想同样深邃，但真正说他是一个物理思想的集大成者还是因为他改革了牛顿的僵化的时空观。

薛定谔对爱因斯坦充满了景仰之情，但他的兴趣并不在广义相对论上，因为广义相对论的数学太难了，高斯他们发明的微分几何他没有学习过。薛定谔在1926年面对的一个困难是来自玻尔集团对他的理论的攻

击和非难。因此，他很想知道爱因斯坦是不是会支持自己的波动方程，他也想知道，那个海森堡的所谓矩阵力学，到底是怎样一个力学。量子力学的一切似乎在一个梦幻之中——他希望能得到爱因斯坦的支持。

9 犀利哥

(1)

在维也纳读中学的时候，泡利就已经是一个非常强悍的少年了，可惜奥地利没有什么好大学，于是，泡利就来到了德国的慕尼黑大学，师从物理学家索末菲。在引力场内被发现弯曲光线的时候，泡利这个21岁少年，就开始写了一系列关于相对论的文章。泡利少年成材，春风得意，他自视极高，对别人也就相当刻薄，可以称为“犀利哥”。



泡利

索末菲是一名数学能力还不错的物理学家，他曾经算一个很难的积分，用到了复变函数的内容，他的原子模型和玻尔的如出一辙。只不过，在玻尔那里，电子绕原子运动的轨道是一个标准的圆，而索末菲觉得，既然在太阳系中，地球等各个行星的轨道是一个椭圆，这是开普勒当年就已经说明了的事情。所以，电子绕原子转动的轨道，应该是椭圆。

椭圆轨道可以修正一系列的光谱项，但这不能完全与实验对应上，于是，索末菲一筹莫展，在此之后，他的研究小组里新来了一个学生，是一个大脑袋方脸孔的矮个子小伙，这就是泡利。泡利长得天庭饱满地角隆丰，一看就是一个聪明的孩子，其发言大胆辛辣，在组内相当活跃，其批判之犀利，可以说达到目空一切的程度，他甚至声称爱因斯坦看来也不是真的那么愚蠢。

泡利因为在高中阶段已经自学完爱因斯坦的狭义相对论，所以他几乎天天给索末菲讲狭义相对论的那点破事，仿佛他才是索末菲的老师。索末菲也受到触动，灵机一动，于是在光谱修正的时候，又加上了狭义相对论修正。这样一来，居然天衣无缝，理论结果与实验结果意外地符合，这让索末菲大喜过望，也对泡利情有独钟。所以这师徒两人，一个搞椭圆一个搞狭义相对论，可以说是珠联璧合。

实际上，索末菲的模型是错误的——正确的模型应该是电子在原子内的运动根本不存在所谓轨道，但是，因为能解释试验，所以人们对他的评价变得很高，玻尔的古典量子理论，也被改称为“玻尔-索末菲理论”。

在狭义相对论的修正中，索末菲与泡利引进了一个很重要的无量纲的纯数字，因为带量纲的数字都不表示物理本质，比如1米的长度，我们可以改称其为100厘米，可以重新定义1米的标准，称之为19米。这样都可以，所以有量纲的物理量只有相对的意义，于是物理学家需要寻找没有量纲的而且在物理理论中自然出现的纯数，索末菲引进的这个数字就是物理学界最重要的常数之一，号称精细结构常数——因为这个常数可以解释光谱的精细结构。精细结构常数的直观的意义是：在氢原子中，电子的绕核速度等于光速的 $1/137$ 。

在这个迷茫的治学过程之中，索末菲还发现，在他的狭义相对论修正之中，如果原子的核电荷数字超过137，那么，他的狭义相对论修正就会失败，这似乎在冥冥之中暗示，原子的核电荷数不可以太大。

至今确实是没发现核电荷数超过137的原子。

所以，这又有点诡异了。

索末菲是老派的物理学领袖之一，当时在欧洲大陆有三大物理学中心，分别由索末菲、波恩和玻尔执掌。爱因斯坦则到处流窜，从苏黎士奔到布拉格，又奔到柏林，他与这三大门派的交往并不深，看上去好像一个独行侠。

(2)

泡利跟着索末菲一起做物理，耳濡目染的就是这些椭圆轨道和狭义相对论修正，因此，他是在很年轻的时候就知道了电子绕着原子是在一个平面上运动的，而且具有椭圆的轨道。这养成了他对物理学的嗅觉，他很喜欢椭圆轨道，等他读了几年大学以后，他对椭圆轨道已经异常了解，他甚至知道，这椭圆的背后有一个隐藏的对称性。

什么是椭圆？在数学上，椭圆的定义是在平面上到两个定点之间的距离之和等于定长的点所组成的集合。这个是很清楚的，一般高中生就要学会怎么样画一个椭圆。

一般说来，一个椭圆是封闭的，这样的对称性背后，包含着守恒的物理量。由对称性导致守恒量，是同时代伟大的德国女数学家诺特（Noether）的思想，数学家外尔曾经这样开玩笑：“女数学家有两种，一种不是女的，一种不是数学家。”没有问题，诺特肯定是一个数学家，她一辈子没有结婚，把全部精力投身给了近世代数。描述对称性最好的语言是群论。对称性和守恒量有一一对应的关系，这一点，是深刻的。比如，众所周知的结论是，空间是均匀的，所以动量守恒。于是，行星运动的轨道是封闭的椭圆，这样的对称性导致的守恒量就是拉普拉斯-龙格-楞次矢量。这个矢量是指向椭圆长轴的方向的。

泡利看着他老师在纸上和黑板上画的那些椭圆轨道出神，他觉得这

背后有很神秘的东西还没有被挖掘出来。在几年后的1928年，犀利哥终于有所领悟，他可以看到这些额外的守恒量正是他需要的李代数，后来他拿这个守恒量子解答出来了氢原子的能级。这是后话。

(3)

前面已经说到，1919年相对论的革命热潮开始感染了一些物理学家，比如爱丁顿和外尔等人就开始写了一些介绍广义相对论的教材，这些书多数被冠以时间、物质和引力这样的名字，其主要强调物质对引力场的决定作用。也就在这个时候，当时有一个百科全书，也想要介绍最先进的相对论知识，于是，编辑就找到了索末菲，请他来写一系列文章。索末菲自己对广义相对论是非常不清楚的，于是把这个任务交给了刚刚上大学2年級的泡利。

广义相对论是以理论晦涩难懂且对数学程度要求高著称，泡利一边自学一边写，比当时我们中国的徐志摩靠谱不少。徐志摩当年在欧洲游学的时候，他的大舅子也就是张幼仪的哥哥张君勱曾经给他一本相对论的书叫他去看，他看了以后囫圇吞枣，最后也写过一篇科普相对论的文章《安斯坦相对主义——物理界大革命》发表在梁启超主办的《改造》杂志上。但可惜的是，徐志摩本身没有搞懂相对论，所以他的科普文章写得一塌糊涂，科普效果很差。而同样差不多年纪的泡利则不同，他的文章不仅总结了当时已有的成果，并且给出了自己的解释和看法——一看就是一个青年物理学家的手法，不像徐志摩属于纯扯淡的。他写的这系列条目后来出版了单行本《相对论原理》。此书出版的时候，泡利还不到26岁。此书有中译本，由凌德洪、周万生翻译。

泡利读大学的时代，是相对论和量子理论两棵小树苗在渐渐长成参天大树的时代，是一个激情澎湃的时代。量子理论有一些基本的问题还没有解决，但留下了浩如烟海的光谱数据。从这些数据里能读出物理的

人都堪称开普勒式的人物。索末菲就是为了解释这些光谱数据引进了椭圆修正和狭义相对论修正。但泡利不满足于解释单个原子，他希望解释门捷列夫的元素周期表。

1922年，泡利离开慕尼黑大学，来到哥廷根大学和玻恩一起做研究。这一年，正好玻尔来讲学，于是，泡利也就结识了这个古典量子论的领袖人物。这年秋季泡利跟玻尔到哥本哈根大学理论物理学研究所工作。

当时物理学家已经在光谱中推测，每一个轨道上的电子个数是有限的，这很奇怪，因为，按照物理学中的势能最小原则，一个跳楼的人总要落到地面，因此在原子中，电子似乎也应该全部集中在势能最小的那个轨道上。但观测的结果表明，每条轨道上，只有有限个电子。

准确地说，每条轨道上，一般是有2个电子。

为什么是2个？

这个问题吸引了很多年轻的物理学工作者，很多人提出了一个解释，认为电子在作椭圆轨道的运动的同时，还有另外一种看不见的神秘的内在的运动，这种运动被称为自旋。但要确证自旋的存在，很多事情纠缠在一起，使得问题无法彻底澄清——主要是1922年开始当时做实验的施特恩-格拉赫（Stern-Gerlach）等人把银原子束在磁场中的偏转成上下对称的强度相等的两束这个经典实验的物理解释给解释错了——他们的实验使用的是银原子（Ag），银原子的电子结构是：2，8，18，18，1。除去最外层电子外，其他为满壳层。相比较于电子的磁矩，原子的磁矩可以忽略不计。因此只需考虑电子运动导致的磁矩，而除最外层5s电子外，其他电子轨道角动量、自旋角动量恰好完全抵消。又因为s电子轨道角动量为零，因此银原子磁矩近似而言主要是由5s电子的自旋导致。当时斯特恩和盖拉赫使一束银原子通过非均匀的磁场，发现银原子分裂为两束。原因是原子磁矩在非均匀磁场中受到力的作用。这实验同时说明，银原子磁矩有两种取值，可惜的是当时人们并没有自旋的

概念，根据轨道角动量的理论，轨道角动量（ L ）的取值只能是整数，如：0, 1, 2..., 磁量子数取值则有 $2L+1$ 种可能性，即由 $-L, -L+1, \dots, L-1, L$ 。因此轨道角动量概念只能解释奇数条条纹分裂，而无法解释偶数条条纹分裂。于是他们就陷入了迷茫——因为他们是为了验证索莫菲的角动量量子化而来进行这个实验的，而银原子的轨道角动量正好是零，因此实验结论与想要求证的东西可以说正好南辕北辙。后来他们用氢原子做了同样的实验，发现还是这样的结果，于是开始想着要引进其他量子数。解决方案是引入电子自旋（ s ），自旋取值为 $1/2$ ，自旋投影取值为 $-1/2, 1/2$ ，正好可以分裂为2条。

但当时施特恩-格拉赫根本没有想到自己是发现了自旋，所以事情变得很复杂。而同时在碱金属元素的发射光谱课题中，在1924年，泡利首先引入他称为是“双值量子自由度”（two-valued quantum degree of freedom）的概念。这使他可以形式化地表述泡利不相容原理，即没有两个电子可以在同一时间共享相同的量子态。但到底什么是他所谓的量子态，他本人也并不清晰。

到了1925年，克罗尼格（Ralph Kronig），朗德的一位助手，提出这个量子态中还有一个叫自旋的量子数，这是由电子的自转产生的。当泡利听到这个想法时，他作为犀利哥马上给予严厉的批驳，他指出为了产生足够的自转角动量，电子的假想表面必须以超过光速运动，而这违反他最熟悉的爱因斯坦狭义相对论。由于泡利的批评，克罗尼格决定不发表他的想法。当年秋天，两个年轻的荷兰物理学家产生了同样的想法，这两个年轻的研究生就是乌仑贝克和古德斯密特，他们在保罗·埃伦费斯特的建议下，以一个小篇幅发表了他们的结果。

尽管泡利最初是反对这个自转的想法的，但他后来发现自转与自旋并不是一回事，电子的自旋是没有经典对应的，因为电子没有大小，电子不像一个地球那么大可以自转，电子没有大小尺寸，所以它的自旋就是一个纯粹的量子现象。于是泡利在1927年形式化了自旋理论，他开拓

性地使用泡利矩阵作为一个自旋算子的表述，并且引入了二元旋量波函数。也就是说，在对电子自旋的数学化的描述上，他终于取得了成功。

这样他也就能大致地说明他的不相容原理了。在每个轨道上存在2个电子，这就好像在高速公路上，每个车道上只开2辆车，这背后肯定有事，说明交通管制了，可能有政府高官正去机场接客，所以车道被管制。

在原子世界，也存在同样的交通管制——这就是泡利在1925年写出的不相容原理：“在同一个原子中没有两个电子可以在相同的量子态。”这意味着，没有两个电子的量子态相同：能量，角动量的大小，角动量的方向，以及最重要的“自旋”。

如果把原子比作一个家庭，那么这个家庭里所有的孩子，都不可能既有相同的年龄，又有相同的相貌，而且还有相同的DNA遗传信息。每一个轨道最多有2个电子，就好像说这个大家族里的每个家庭最多有2个孩子，而且必须一个是男孩，一个是女孩。

原子就是这样一个巨大的空间，里面有很神秘的运转法则，泡利的不相容原理，其背后的思想核心是电子存在一个额外的自由度，那就是自旋，为了说明这个非相对论性的自旋，泡利引进了自旋角动量SU(2)李代数。关于自旋的这个不相容原理是年轻的犀利哥做出的第一个物理学猜想，虽然他无法从更深邃的角度来证明这个猜想是对的。但很多人还是把这个猜想称为一个原理，这是从光谱数据中得到的第一个宝贵原理。这也是弗朗禾费时代以后，光谱学研究的最高成绩之一。时过境迁之后，人们最后还是把泡利不相容原理应用到了恒星研究，也得到了白矮星的质量上限。

在泡利的不相容原理提出来的同时，犀利哥有一个师弟在光谱中得到了另外一些重要的东西。

10 海森堡：日出

(1)

1923年夏天一个闷热的傍晚，云层在高空形成湍流，校园里没有一丝风，热浪让空气无语凝噎。德国慕尼黑大学的一间教室坐着几个教授模样的人，他们已经听完了一个23岁年轻人的博士论文答辩。

其中一个教授模样的人跷着二郎腿，不屑地对年轻人说：“海森堡，你这样下去是不行的啊——虽然我听说你父亲在这个大学里也是有名望的教授.....但是，我们是要讲事实的，你的湍流理论缺乏实验数据的支持，模型也有问题，我在考虑要不要给你这个博士学位。”

年轻人低着头羞涩地站在讲台边上，唯唯诺诺地说：“维恩教授，其实，关于湍流，我.....”名叫维恩的教授粗暴地打断了年轻人的话，说：“别再解释，你的论文不行，你实验也不行，理论也不行，你这样混一个博士学位，对你个人的成长也很不利.....”

年轻人似乎有点愤怒了，他的眼睛里充满委屈的愤怒光芒，血液像岩浆一样在沸腾，拳头已经紧紧地攥起来。心想小爷我15岁自学微积分，高中毕业读完外尔写的相对论.....海森堡听见自己的喉结在颤动，吐出这样的一句话来：“您说得对，维恩教授。”

接下来，就是一次长时间的比较空洞的沉默。

教室里静得连一根针掉在地上的声音都能听见，慢慢地海森堡也好像中了化骨绵掌，虚汗从后背冒出来，仿佛一滴一滴掉到了地上。海森堡的导师索莫菲终于打破了这短暂的死寂，说：“维恩教授，其实，海森堡对实验技术确实缺少了解，但.....依我看，他的文章还是有可取之处的.....”

海森堡感觉自己好像一个可怜虫，静静地等待着命运的审判。维恩

的刁难成了他心中的刺。维恩当时已经得到诺贝尔奖，他对黑体辐射有一定的研究，得到了一个维恩位移定理，说黑体辐射最强的波长和温度之乘积是一个常数——类似于汽车在额定功率一定的时候，汽车引擎产生的拉力和汽车的速度之乘积是一个常数。

海森堡得到博士学位以后，连夜离开了慕尼黑前往哥廷根，23岁的脸上还充满稚气，但这次博士论文答辩已经让他成熟了不少——他的人生观已经悄悄改变了，他变了，内心深处变得怨愤——没有人知道他已经长大了，心中充满了不平。他前去投靠玻恩，这事情是早已经说好了的，1922年10月他们已经认识，海森堡这次去玻恩那里相当于是去那里做博士后研究——人生若只如初见，交往越多，关系越微妙。自从投靠上玻恩以后，海森堡也渐渐地讨厌起玻恩来，到了最后，他视玻恩为寇仇，在文章里也很少提起玻恩对他的影响和鼓励，甚至在1932年诺贝尔演讲中他似乎也对“玻恩”这个人名讳莫如深——玻恩实在很委屈，他心里说：“海森堡，我和你前世无怨今生无仇，你怎么凡事都要刻意冷落我。”

人是很微妙的动物。

海森堡在玻恩那里开始他新的工作。1924年复活节，他第一次去哥本哈根，但不久就回到了哥廷根。这个时候，表面上看上去有点腼腆的青年海森堡正在积蓄足够多的力量——在他心中，那个曾经的阳光大男孩，已经死去，现在活着的，是一个内心世界如静水深流的海森堡。海森堡不由得回忆起自己的青少年时代……

19世纪末到20世纪初的电力革命让德国走在了世界最前沿，以西门子为代表的电器公司增强了德国的科技与经济实力，渐渐成为后起之秀，这个国家在很多方面已超越英法等老派帝国主义国家。可惜在俾斯麦统一德意志的时候，地球上的殖民地已经被西班牙、葡萄牙、荷兰、法国，尤其是大英帝国给瓜分干净了——那个大英帝国不是曾经号称是“日不落帝国”吗？德国只在非洲几个贫瘠的地方扶植了自己的势力。

要打破前人的格局，德国联合了奥匈帝国与奥斯曼帝国组成同盟国，开始与英法对抗。英法自然不敢小视，拉拢了和德国几乎同时崛起的美国与俄罗斯组成协约国。第一次世界大战在巴尔干半岛这个火药桶的引爆下展开了。



绘画：张京

从1914年“一战”的正式开始，德国大量的民用资源被消耗在无意义的战场上，国民经济开始崩溃，人民的生活也变得拮据。有一次海森堡因为太饿了，直接从自行车上摔了下来。众所周知，一战之后，德国战败，凡尔赛和约签署，德国把阿尔萨斯与洛林地区归还给法国，同时要承担巨额赔款，在这个时期成长起来的德国年轻人进入了食不果腹的贫穷时代。笼罩在战争失败的阴影下，慕尼黑的街头天天上演着几个政治派别的血腥搏杀，弱肉强食成为社会常态。海森堡切切实实地感受到了政治的虚伪和残忍，也许只有在自然科学才能找到那份纯真与善良。成王败寇的丛林现实让人感觉到只有让自己变得更出色，更强大才能顽强地活下来.....

1925年5月，天空那么阴。

北海。

赫尔兰岛。

海鸥在盘旋。

孤岛，杂草丛生。

海森堡戴着墨镜，脸色阴郁，走在沙滩上，用脚趾踢打着浪花。他得了枯叶草病，是一种花粉过敏的病，需要在这一个没人的地方躲一段时间。

海浪打在沙滩上，发出哗哗的响声，那么有节律，这个单调节律在海森堡听起来是一个周期运动，他的脑子里还在想着另外一个周期运动，那就是电子绕着原子核的圆周运动——这是玻尔的模型，他已经厌烦了。

24岁的他，厌烦了一切，他喜欢海浪的声音，那像是一个孤独的沉默者的哭泣，拍打着他的灵魂，他爱大海的蔚蓝和深邃，他知道海浪里还有他所研究的湍流，但现在他更关心一些简单和谐的东西，那就是月球绕着地球转，潮起潮落这些简单的周期运动。玻尔的模型那么单调，简直有些无聊，因为电子的圆周运动的轨道根本是看不到的。只有光的频率和强度，才是可观测的。24岁，嫩得像一棵草，他决定出手了——干掉玻尔！

电子轨道？周期运动？

电子轨道是周期性的？

周期函数可以展开为傅里叶级数？

展开它？

展开以后？是频率和振幅？有意思，值得一搞？

哦，行，就这样干，海森堡像被沙滩上的什么东西吸引住了，蹲下来，捡起了一颗美丽的鹅卵石，他将之攥到手心里，然后缓慢地站起来，用力将之抛向远处的大海。

从海边回到他住的旅舍，洗了一个澡，准备把刚才脑子里想到的那个东西写下来。这时候，已经是晚上了。

等海森堡连夜写完这个傅里叶级数展开，他发现，这个傅里叶级数不应该使用正常的傅里叶级数，因为原子发光的时候，光的很多频率并不是等间隔分布的，光谱线的频率之间基本上显得杂乱无章，但这些频率也可以作为一种变异的傅里叶级数的展开频率——这就好像人民币一样，不是1元，2元，3元，4元，5元，6元……面值为等间隔的纸币都有，而是只有1元，5元，10元……这些基本面值，但同样可以用这些面值去展开任何需要支付的钱款数目。如果这种变异的傅里叶级数展开是可行的，那么两个轨道的乘积满足一个很奇怪的求和规律。这是什么呀？海森堡写到这里，就停住了，陷入迷惘之中。他觉得自己像一个迷路了的旅人，这篇文章意味着什么呢？

写完文章后，已经是凌晨，困意全无，海森堡出门，跑到远处的山崖上，静等旭日的升起。

天边首先露出一片鱼肚白。

海森堡的思绪仍然萦绕在那些令人困惑的量子问题上，现在他压抑已久的灵感不断迸发，潜在的革命性思想正在努力冲破传统观念的束缚。他在想，既然电子没有轨道，那么通常的位置和速度描述将不再有意义，这样看来，确实必须利用新的描述量来建立理论。玻尔的对应原理、克拉默斯的色散关系不断出现在他的脑海中，他意识自己所创造的这种变异的傅里叶级数展开也许确实可以建立一种新的力学理论。

海森堡从北海回来，把文章交给玻恩阅读，问道：“这篇文章值得发表吗？”

玻恩的回答非常言简意赅，只有两个字：“值得。”

过了一会，金语良言的玻恩又补充道：“那些乘积就是矩阵，我上学的时候就跟希尔伯特他们学过，应该就是矩阵。”

于是，海森堡就发表了他一个人署名的文章，第一篇矩阵力学的论

文《关于运动学和动力学的量子力学解释》，历史上称为“一人文章”。这个文章里有很奇怪的乘法求和法则，玻恩一语道破天机，认定这个乘法求和规则，正是英国数学家凯莱所定义的矩阵乘法。

凯莱先定义了矩阵的零元素和单位矩阵，接着通过坐标系的连续两次变换，发现了一个很自然的矩阵乘法的定义。

量子力学不是一门直观的物理理论，但这个理论具有最直观的表达方式，那就是矩阵。矩阵在生活中经常出现，如果在一个大学里，夜晚站在女生宿舍楼下，就能看到矩阵——虽然矩阵的英语matrix亦有“子宫”之意，不过此处我们并不研究子宫，而是谈论矩阵。



宿舍楼每一个窗户里，住的人的数目构成一个宿舍矩阵的矩阵元

如果盯住整幢宿舍楼看，假设这个楼是8层的，每层有8个宿舍。那么，这就是一个8乘8的矩阵，而每一个宿舍里女孩子的人数，就可以看成是矩阵元。矩阵是多么美妙。文科读者们一定不要有惧怕心理，其实真理永远是朴素的，矩阵就是一堆数放在一堆整齐的方格里而已，欧拉以前也没有搞过矩阵的乘法，但他曾经思考过一个问题，这个问题是所谓拉丁方问题，或者说“三十六军官问题”——本书不再展开谈这个，有

感兴趣的读者自己可以上网查阅。总之，欧拉为了把一些数字放在一起做成一个满足某种性质的矩阵，花了九牛二虎之力。其实早在中国古代，传说大禹治水时（约公元前22世纪），于洛水中浮现一只神龟，它的背上有规律地排着九种花纹，这图后人称为“洛书”，也称九宫图。如果把图形改成现在通行的阿拉伯数字，就是一个3阶魔方（Magic square）。

$$\begin{array}{ccc} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{array}$$

上面这个魔方可以具有非凡的特性，它其实也是一个矩阵。如果把它看成矩阵，那么也可以想到计算它的特征向量，做一系列的操作。也可以把它们乘起来，研究其他的性质。因此确实可以把魔方看成是矩阵的一种——作者就曾经发现 n 阶魔方 M 与其转置矩阵的乘积的最大本征值等于 M 的迹，这是一个很让人惊讶的结论。如果您还是不懂什么是矩阵，请先不要着急，随便找一本线性代数的书就可以学会，为了激励有志青年学习矩阵理论，我们来看哈代如何评价华罗庚，哈代说“华玩弄矩阵就好像玩弄整数一样轻松”。

要想轻松一些，我们可以来看一个最简单的矩阵的例子。比如某一天，从北京开往杭州的火车，在停靠站的时候，上下车的人的数目如下图所示（从矩阵中可以看出北京到济南站的时候，有100人下车，有48人上车）。

火车T31/T32上下车人数表

下车 上车	北 京	济 南	南 京	上 海	杭 州
北 京	0	100	300	400	1200
济 南	48	0	55	45	33
南 京	99	122	0	49	77
上 海	49	132	97	0	100
杭 州	1100	48	88	151	0

(3)

海森堡是在非常懵懂的情形之下，发现如果采用直接可以观测到的光谱频率作为基本可观测的物理量，而抛弃不可观测的电子轨道，那么在新理论（量子力学）里，一些物理量应该用矩阵来描述，而不是以前认为的函数或者数字。这是量子力学全部的数学意义所在。

而一个n阶矩阵M有特征方程

$$f(x) = \det[M - IX] = 0$$

这个特征方程是一个n次多项式方程 $f(x) = 0$ ，在第1章我们已经讲过，代数基本定理说，n次多项式方程具有n个解，这n个解被称为矩阵M的特征值。在很多情景下，当M表示一个物理系统的能量时，上面所讲的n个解正是这个物理系统的能级。

凯莱当年还发现一个更加有意思的事情，那就是上述n次多项式方程 $f(x) = 0$ ，对于矩阵M也是成立的，也就是说， $f(M) = 0$ 。这被称为凯莱-哈密顿定理。

量子力学的基本语言是矩阵，这起源于玻恩的贡献。玻恩对薛定谔

的波函数和海森堡的乘法求和规则做出了正确的解释，所以，他是量子力学历史上，在正确的时间正确的地点出现的最正确的人。

海森堡后来说：“那是在夜里三点钟左右，计算的最后结果出现在我的面前，……我深深震惊了。”1925年的漫漫长夜之中，当别人还在对电子轨道恋恋不舍、犹豫不决时，彻底抛弃那些看不见轨道的海森堡终于发明了一套新数学方案——变异的傅里叶级数展开，它会导致一种新的犹如九九乘法表的东西，不过这个参与乘法的已经不是整数，而是矩阵了。当时矩阵对物理学家来说，还是很神秘的东西，海森堡被数学和物理夹杂，感觉如有神喻，那天晚上怎么也睡不着，他早早起床，站在海边的礁石上，一个人静静等待旭日的升起。

外一篇 测量问题举例：对波长的测量

(1)

通过阅读本书，读者们一定已经发现，这本书已经提到过关于一个物理量到底是不是可以测量的这些技术性问题。其实，量子力学理论是非常强调“观测”的。这也是物理学和数学的区别——比如一只苍蝇绕着一个坐在凳子上的人转圈圈，在人看来，是苍蝇在动，而在苍蝇看来，是人在动，那么到底是谁在运动呢？这是一个在相对论里面最基本的问题，通过一般的哲学思辨是无法找到正确答案的，只有通过测量苍蝇与人的四维加速度，才可以从物理上得到答案——关于相对论的测量部分，我们将在本书的第二部分进行讲述。

在这一章，已经讲到，哥本哈根学派和海森堡他们手中有一把利剑，这把利剑正是“可观测量”，可观测量的物理量其实是要对应于量子力学中的厄米算子——你能测量到很多不同的结果，是因为这个厄米矩阵有很多不同的特征值。所谓算子，就是“矩阵”，矩阵的英文是matrix，

电影《黑客帝国》的英文名字就是matrix。有一种数学软件叫Matlab，就是matrix Lab的缩写，意思为“矩阵实验室”。这个数学软件的功能非常强大，矩阵的功能原本就非常强大——几乎能解答从微分方程到数论的大部分数学问题。

我们暂时不谈矩阵，虽然海森堡的力学，也被称为矩阵力学——我们要来看看到底怎么样具体地测量一个与量子力学相关的物理量，比如光的波长，到底是怎样测量出来的。

(2)

任何仪器都是人的身体功能的延伸，比如电视机可以帮助我们看到空中的电磁波激发出来的电视图像。人同样可以被看成是一台仪器，这台仪器也有很大的局限性，比如，你的眼睛是不能直接看出一束激光的波长的——虽然你可以看到激光笔射出的红光，但不能看出其波长。

为了探测一束光的波长，需要一台光谱仪器。从仪器学的角度来看（仪器学不是一门公认的成熟学问，这门学问包括光学，机械，电子等多学科门类，是一个综合性学科，著者在这里冒昧地先引进“仪器学”这个名词），探测光的波长，至少需要3个组成部分。

1. 光源
2. 单色器
3. 检测器

我们将依次来泛泛地谈论这三个部分，从而让读者管窥其中的意义。

(3)

光源相对比较简单，任何会发光的物体都可以成为光源，在初中物理课本中，一般分为自然光源与非自然光源，总之比如太阳就是自然光

源，而电灯则是非自然光源。一般来说，如果我们使用一个电光源的话，一个钨灯就是一个典型的例子。这背后有非常深刻的物理，你无法通过理论计算电功率而确定出钨灯发出的光的整个光谱，因为这不是简单的黑体辐射。钨灯发出的光谱是连续的。还有一些灯能发出很尖锐的脉冲峰，比如空心阴极灯。

空心阴极灯又叫元素灯，它能发出一些特定元素的特征谱线。但因为谱线不可能是无限细线光谱（存在碰撞展宽，多普勒展宽和自然宽度），所以，其实空心阴极灯发出的光谱也是在波长上连续分布的。谱线具有宽度就是真正的物理学，只有在数学里我们才可以谈论无限狭窄的线光谱——基本可以肯定的是，任何涉及无限的数量都不是物理学中的物理量。谱线的展宽具有很多类型，比如自然宽度起源于海森堡的不确定原理，多普勒展宽则起源于发光原子的热运动……鉴于本章只想谈论测量光谱线的波长，我们只需要记得一点，那就是，任何光源发出的光谱线，都是有一定的宽度的，没有百分百纯净的单色光——那些理想化的状态，只能让那些数学家与哲学家去研究，物理学家必须面对现实。

(4)

单色器的主要作用是把一个连续光谱的光从频率上分离开，也就是说，输入单色器的是一个复合光，输出单色器的是一个单色光（理论上的单色光，实际上单色光是不存在的，这是仪器原理决定的，也是真正的物理）。

一个单色器最简单的结构是入射狭缝、光栅和出射狭缝。狭缝的宽度是决定光谱带宽的，而光栅则起到分光的作用。在目前，比较多用的是闪耀光栅，这种光栅可以使得出射光能量不集中在零级，而是集中在我们需要的波长附近。当然，对于闪耀光栅的出射能量随波长的分布，

有很多种计算方式，最近比较流行的，是采用傅里叶光学的观点。

仪器为了实现光栅方程，需要一些机械结构。因为一般来说，光栅转角和出射波长之间存在正弦关系，这个正弦函数需要机械结构来实现，所以肯定会有误差。

一定要记住，无论在哪个时代，有些事情一定是不可能做到的，因为机械误差的存在，加上物理学原理本身的限制，单色器发出的光不可能是真正的单色光。这就是“不确定关系”最直观的意思，因为仪器永远测量不到物理量的真实值，大家就想到用多次测量求平均的方法来逼近真实值，这在概率论上被称为“大数定律”，其意思是说，只要你测量的次数足够多，你得到的数据的平均值就会很接近真实值。从某种意义上来说，这里面肯定有逻辑的缺陷，换用量子力学的思考方式，我们最好干脆一些，其实“真实值”并不存在！

因为你永远测量不到“真实值”！（对于比较脆弱的量子力学系统，测量仪器的介入会影响这个系统本身的行为。）

（5）

检测器是把光信号转变为电信号的传感器，一般叫做光电池或者光电倍增管。最简单的实现方式是基于爱因斯坦光电效应原理的光电倍增管。如果仅仅谈论对波长的探测，检测器的知识非常简单，它总能把光信号转化为电信号。但从工作过程的细节来看，这个部件涉及电流噪声以及后续放大电路的设计，以及最后的模数转换（把模拟量转化为数字量，非常像从连续的经典力学进入离散的量子力学）。

我们不再絮叨，展开来讲，仪器的整个工作过程涉及能量的传递和信噪比的传递，很多细节都可以单独写出一本书。但就噪声而言，仔细分析就会知道，在傅里叶分析的意义上，噪声是有颜色的——人们把那些完全无规律的噪声叫做白色噪声。

11 交换游戏

(1)

命运女神，有时候会把很多不同的人物的命运安排得非常奇妙，“天遥地远，人生何处不相逢”。因为海森堡下面的这次旅行，把一个电机工程师也卷了进来。

海森堡把“一人文章”发表以后，就离开欧洲大陆去英国访问，他到剑桥的时候，见到了福勒（William Alfred Fowler）教授。

海森堡说：“福勒教授，我写了一篇文章，想给您看一下。”

福勒说：“行，你先给我吧，我会看的。”

福勒把海森堡的论文交给了自己的一个研究生去看，这个研究生是一个沉默寡言的小胡子，一个三棍子打不出一个闷屁的电机工程师——此人本科是读电机的，可以说此人本非池中物，一遇风雨便化龙，这正是后来居上的狄拉克，他不是物理学的科班出身，也就晚了大概半年就加入到这个物理学大游戏当中，但是，先胖不是胖，后胖压塌炕，狄拉克出现以后，手法之凌厉娴熟，对物理之洞若观火，让人目瞪口呆。

刚开始狄拉克没有怎么仔细读海森堡的“一人文章”，似乎兴趣不是很大——这当然也要怪海森堡本人写的文章实在是玄虚晦涩，一般人读他的文章根本找不到他的思路，乱糟糟的好像一团乱麻堆在这里，这就好像是一锅东北乱炖，不符合狄拉克的清淡口味——狄拉克自己写文章一向言简意赅，逻辑条理十分清楚，被称为“秋水文章不染尘”。

狄拉克要等到他看到约当与玻恩合写的所谓“二人文章”，才仿佛长了飞毛腿迅速赶上。

事情显得很诡异——读者们会看到，这个电机工程师具有异常深刻的数学功力——电机工程师有数学才情的这类珍稀物种在历史上相当罕

见。在他之前只有一个叫Heavyside的电机工程师表现相当抢眼，是小半个数学大师，他发现，微分算子和积分算子可以看成是互为倒数。所以这个人解电路微分方程，做起来就像解代数方程——这后来被证明是正确的，其实就是拉普拉斯变换。

那么，“二人文章”又是怎么回事呢？原来海森堡去英国访问的时候，他的博士后老板玻恩找了一个22岁的年轻人来干活。这个年轻人就是约当。玻恩为什么要找一个嘴上无毛的毛头小伙子来呢？原来，玻恩虽然已经看出海森堡文章是在使用矩阵的乘法，可是，和现在的很多研究生导师一样，他只能在宏观上把握事情，细节上很困难，于是找了一个很懂矩阵的年轻人一起来另外写一篇文章介绍量子力学的思想，这就是《论量子力学》，这篇文章就是历史上著名的“二人文章”。

这个时候，其实已经埋藏下危机——很多年以后，约当说，当年这个署名玻恩和约当的“二人文章”——其实出自他一人之手。（读过研究生的同学们一定是深有体会的，有些文章导师是不参与写作的，只需要把名字挂上去就可以了。）问题可能更微妙了，因为海森堡根本不知道他们两人要写这个文章，而很明显这个矩阵力学的思想是自己先搞出来的，现在他们把自己撇开写了一个“二人文章”，这让海森堡有所不快。

当时他们两个人的文章，第一部分对物理学家来说比较新鲜，就是介绍矩阵的乘法，比如如下乘法：

$$\begin{vmatrix} 4 & 2 \\ 3 & 1 \end{vmatrix} \times \begin{vmatrix} 2 & 3 \\ 4 & 1 \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

怎么求出a、b、c、d这四个未知数？当时的物理学家对这套还很陌生，但约当很熟悉，三下五除二就搞定了。



作者张轩中所开发的矩阵乘法计算软件，图片中的2个男子为狄拉克和海森堡

(2)

英国数学家凯莱早已经定义了矩阵乘法，因此上述矩阵计算自然是水到渠成。并且有一个有趣的性质，那就是在一般情景下，两个矩阵A和B的乘法具有不可交换的性质，也就是说

$$A \times B \neq B \times A$$

打个比喻来说，先谈恋爱再结婚，与先结婚再谈恋爱，一般来说，结果并不一样。对比上图片的计算结果，以下图片说明了矩阵乘积的不可交换性。



海森堡以及玻恩他们已经看到，经典力学的很多物理量，如果要改写到量子力学里，都要表示成矩阵，那么传统的动量 p 和位置 q 这两个物理变量，现在也成了矩阵，而且，它们并不遵守传统的乘法交换率：

$p \times q \neq q \times p$ 。

这个时候，应用到单粒子的物理动量与坐标，波恩和约当把 $p \times q$ 和 $q \times p$ 之间的差值也搞了出来，结果是这样的：

$$pq - qp = (h/2\pi i)I$$

两人写出来以后，感觉量子力学春光乍泻。但很明显， p 和 q 不可能同时是有限大小的矩阵，因为对于两个有限行有限列的矩阵 p 与 q ，乘积 pq 与乘积 qp 具有相同的对角元。

也就是对有限大小的矩阵的“迹”：

$$\text{tr}(pq - qp) = 0$$

这显然与 $pq - qp = (h/2\pi i)I$ 是矛盾的——因为单位矩阵 I 的迹并不等于零。

看出这里面奥秘的人，正是那个年轻小伙子约当，因为他曾经帮助希尔伯特和柯朗编辑过《数学物理方法》，数学才情比玻恩要高一些，

他很清楚地知道， p 与 q 不是简单的有限矩阵，实际上应该是无限大的矩阵。

可是草稿纸那么小，无限大的矩阵写不出来，于是，这两个人愣住了。

1926年的这个时候，玻恩和约当走到哥廷根的街上喝咖啡，每天内心深处有一些隐忧：不可交换的关系 $pq - qp = (h/2\pi i)I$ ， p 与 q 不可能同时表示为有限矩阵。

为了解决这个问题，两人开始了漫长的琢磨，那么，这个神秘的不可交换关系，到底应该怎么表示。

过了没有几天，哥廷根有一个美国人进入了玻恩的视野，玻恩很想拉他过来合作一把。这个美国人不是别人，正是数学家维纳。维纳是历史上著名的神童，他曾经写过一本书，叫《昔日神童》，讲的就是他自己。读者们有兴趣可以查阅这本书，不过著者不准备谈论这些，因为神童其实到处有，只不过多数人“幼有神童之誉，长而无闻，终乃与草木同朽”。但是，维纳实在是一个例外，他一辈子都是天才。虽然他是一名数学家，不过，他和狄拉克一样，也是一个电机工程师，他写过一本书，书名为《控制论》。此书已经在中国大陆出版，有兴趣的读者可以买来一读，会发现他确实是一个电机工程师。所以，现在的量子力学几乎快成了电机工程师们的天下了。

(3)

维纳到哥廷根之前，在剑桥跟哈代做数论，他第一次看到李特伍德（littlewood），说了这样一句话：“没有想到世界上真有李特伍德这个人，我还以为littlewood是哈代为他写的比较差的文章署的笔名呢。”维纳因为少有天才，所以是有点目中无人的，后来他到过中国，在清华大学讲数学课，一开始听的人不少，讲了几天后，听众作鸟兽散，只剩

下一个人——这个人就是华罗庚。维纳看到这群黄皮肤黑头发的东亚病夫之中，居然有如此杰出的人才，于是推荐华罗庚前去剑桥跟哈代做数论，后来中国才出现一个真正的大数学家。第一次世界大战期间，维纳曾在马里兰（Maryland）的阿伯丁实验场（Aberdeen Proving Ground）进行弹道学研究工作，其主要工作是编制数学用表，主要用于确定火炮瞄准飞机的方位。第二次世界大战爆发，要求防空火力对飞机进行更为精确的跟踪，美国军方又一次邀请维纳参与其研究工作。1940年，美军军方请维纳去阿伯丁实验场重新制订火力表。军方说，希特勒的战斗机速度很快，飞行员们有大耍曲线、翻筋斗的伎俩，英法美盟军的高射火炮根本打不着他们。地面指挥官和炮手们为此很苦恼。因此，需要制订新的火力表。尽管维纳是一个激进的和平主义者，但他还是作为电机工程师式的数学家，参与了军方的研究工作。维纳说用手工和计算尺计算肯定不行了，只有研制高速的计算机。维纳特意写信给时任美国总统罗斯福科学顾问的布什，提议制造电子计算机，并提出了制造电子计算机的五点要求，即计算机五原则。布什表示同意，并立即立项，于是，号称世界上第一台电子计算机的ENIAC就这样诞生了。为了打击空中目标，防空火炮在开火的一刹那必须瞄准目标前方某一点，要想提高火炮的命中率，这个提前量必须迅速而准确地计算出来。尽管德国飞行员善于飞行机动，但飞机的飞行是有规律可循的，特别是当飞行员的机动结束时。凭借对预测问题的研究基础，以及在麻省理工学院使用模拟计算方法的经验，维纳提出了将防空火炮与雷达结合使用的方案：雷达获得的目标数据经过数学运算生成瞄准数据并送往炮塔马达，由炮塔马达自动完成瞄准。后来，维纳提出的方案获得了成功。然而，对于这种新的防空火炮自动控制装置，产生了一个分类的问题。人们认为这种自动控制系统属于动力技术领域，而不是通信技术领域。维纳不同意众人的观点。维纳认为，自动控制系统是通信设备。依维纳的观点，在防空火炮自动控制系统中，电动机的作用是向炮塔传送设计参数，因此，可以把

马达和控制马达动作的计算机看作是通信设备。维纳将计算机看作“另一种形式的通信设备”，认为“它与信息而不是动力存在着紧密联系”。特别令维纳兴奋的是，他发现，防空火炮自动控制系统的运动与生物体运动有着惊人的相似之处。这些全是后话了，总之，维纳当时在哥廷根的时候，还很年轻，30岁刚出头，正是青春好年华。

玻恩谄媚地说：“维纳老弟，我有一个数学问题，想跟你一起研究一下，我们俩写一篇文章怎么样？”

维纳不屑地说：“先说说看，什么问题？”

玻恩有点儿着急，神秘地说：“这个问题关系到量子力学的生死。就是不对易关系 $pq - qp = (h/2\pi i)I$ ，p与q不可能是有限矩阵。那么p和q到底应该怎么表示？”

维纳冷笑着说：“既然是关于所谓量子力学，那我们开始吧。不过依我看，你似乎在搞表示论。”

玻恩闻言，一愣，像一个在河里游泳的人脑袋被人踹了一脚，半天说不出话来……

几天以后，维纳就和玻恩写了一篇文章，原来，p和q不能用有限矩阵表示出来，但可以用微分算子来实现。

$$p = -ih \, d/dq$$

熟悉数学的人，很容易检验，上面这个微分形式可以实现不可交换的关系。对于文科读者来说，一定已经晕头转向了，其实打个比喻就是，一开始 $pq - qp = (h/2\pi i)I$ 这个数学关系，p是男人，q是女人，这个不对易关系就是男人和女人的婚姻关系。可是，在有限矩阵中，根本不可能实现这个关系——也就是说，在一个村子里，虽然有男人，也有女人，但因为双方年纪和家庭相差太大，不能实现婚姻关系。要想实现婚姻，必须扩大配对的范围——去别的村子发展适婚对象。

维纳和玻恩实现了这个婚姻关系，成为一对出色的媒人。爱因斯坦

也注意到最新的进展，在给好友贝索的信中对新理论评价道，“近来最有趣的理论成就，就是海森堡-玻恩-约当的量子态的理论。这是一份真正的魔术乘法表，表中用无限的行列式（矩阵）代替了笛卡儿坐标。它是极其巧妙的……”。

外一篇 不确定原理

（1）

在前面我们基本讲了量子力学的一些基本形式：波动方程与矩阵力学，但是量子力学中最广为人知的基本原理是不确定原理。

先再介绍一下时代背景：1921年，爱因斯坦得到诺贝尔物理学奖（但是由于历史的原因，爱因斯坦在1922年才领到这一奖章），1922年，玻尔得到诺贝尔物理学奖，从此在物理学的江湖上有了两个潮流派系的领袖人物。

1922年，德国科学家为了庆祝玻尔获得诺贝尔奖，特地举行“玻尔节”，邀请玻尔到哥廷根来演讲。那时候，慕尼黑的希特勒正在内心深处筹划啤酒馆政变，他活着就是要改变世界。物理教授索末菲领着两头初生牛犊——泡利和海森堡日夜兼程，横穿整个德国，赶到哥廷根来朝拜玻尔大师。于是，在哥廷根玻尔演讲现场的听众里，有一位21岁的大二学生——海森堡。

海森堡那时候虽然是一个无名小卒，但他是有备而来，准备了一些尖锐的问题。比如，玻尔的理论说明电子要么在这个状态，要么在那个状态，中间状态是没有的——这就在好像说，一个人要么在北京，要么在天津，但他是不需要路过廊坊的。这确实有点反直观，所以，年轻人海森堡对玻尔老师的理论，很想吐槽一番。

在听演讲的过程中，海森堡突然站起来，他听见自己的嘴巴突然一

张一合、一字一句地从那里蹦出来下面这几句话：“玻尔老师，相对论说速度是有限的，最大的速度是光速，电子从这个状态跳到那个状态，不管距离多么短，一定是需要时间的。那么在这段时间内电子处于什么状态？”

玻尔冷不丁被打断了演讲思路，脑子一下子缺氧。嘴巴也就跟不上了，回答不了这个毛头小伙子提的问题，但姜是老的辣，他显得很从容，说：“好问题，it is a good question，我们私下再谈。”

接下来，玻尔继续他的演讲，但方寸有点乱了，内心里很喜欢刚才那个打断自己的小伙子，他心想这孩子显得比较成熟。

会后，玻尔邀请海森堡去散步。

这一次散步，使得海森堡认识到，玻尔虽然在江湖上名气很大，但平易近人，而且对物理的洞察力又很深邃。于是，海森堡暗暗下决心，以后要追随玻尔。

这时候的海森堡心灵土壤，埋藏了这粒种子。于是，他算是跟玻尔已经认识了——这是参加学术会议的好处之一，可以认识行业内的大牛人。

此后四年，海森堡自己没有想到，自己居然领导潮流，物理学1925年开始的高潮就是 he 先搞起来的，这高潮持续到第三年，到了1927年。

1927年，大家都很年轻，泡利比海森堡大一岁，海森堡比狄拉克大一岁，这些年轻人中，还有约当、维格纳和冯·诺伊曼。所以，那时候的物理学，被称为“男孩物理学”。

而薛定谔已经40岁了，他这个时候因为搞出了波动力学，算是大器晚成，1926年10月也曾经来到哥本哈根访问，他住在玻尔家里，专门给玻尔讲解波动力学。可惜，玻尔根本不相信波动力学是对的，于是，大家闹得很尴尬，薛定谔住了没有几天，就病了，他相信自己不应该来说服玻尔。

(2)

1927年2月，乍暖还寒，物理学也一样，暖流已经袭来，但冷空气还没有消退。

薛定谔离开哥本哈根以后，玻尔和海森堡继续讨论，他们觉得很费解，电子一会儿表现得像一个粒子，一会儿又表现得像波。电子如果真是波又是粒子，那情何以堪。

薛定谔回到维也纳，觉得玻尔智商不行，他写信给朋友说：“那些人其实都不怎么懂物理，我是这个混乱时代中幸存下来的唯一一个脑子还清爽的人。”

这个二月，大家的精神高度紧张，脑子也快炸了。

玻尔对海森堡说：“海森堡，明天我要带着你师母去挪威大峡谷滑雪，我们需要休息一下了，脑子要爆炸了。——你也可以出去玩玩，找个小姑娘什么的，你也该谈恋爱了……”

海森堡说：“行，希望旅游能让您充满灵感。”

海森堡这个时候突然起了一个怪念头，他感觉道德的苍白，心想，玻尔这次去滑雪，不会摔断了腿再回来吧。

翌日，玻尔出发去了挪威。

海森堡独自一个人留在哥本哈根，没有玻尔在边上，他的思绪很自由。他心想，如果薛定谔的波动方程里出现的波包真可以描述粒子，那就是说，电子的位置是不确定的，那么好，请问，这个电子的位置，总有一个概率分布，这个概率分布的不确定度是多少？

海森堡计算了一下，发现，位置不确定度 Δq 和动量不确定度 Δp 必须满足如下关系

$$\Delta p \Delta q > h/2\pi$$

其实，数学家一看就会明白，这可以从傅里叶变换中直接推出来，但对海森堡来说，这个结果很是惊人，因为这里面有一个不等号。电子

的位置分布和动量分布，不能同时确定？

这想法把海森堡雷到了。

这个结果是那么强，以至于被认为是量子力学中最深刻的东西，因为按照这个原理，1922年海森堡初次见到玻尔时候的那个问题迎刃而解：其实电子从一个状态变到另外一个状态，这中间确实需要时间，这个时间 Δt 与两个状态对应的能量差 ΔE 有关系，同样满足

$$\Delta t > h / (2\pi\Delta E)$$

海森堡在玻尔不在场的房间里，回答了5年前自己提出的这个问题。写字台上的台灯忽明忽暗，似乎正是这一原理在起作用，海森堡觉得万分孤寂，他激动地趴在窗台上，寂寞的研究所被雪花笼罩，陷入一片雾霭之中，他高喊起来：“来人呀，我发现了上帝的秘密。”

12 物理学牛市：矩阵的运动方程

(1)

1925年的冬天，是一个充满希望的冬天。在这个时候，海森堡从英国回到了哥廷根，看到玻恩和约当的文章已经出来，心里有点不快，因为矩阵力学是他开的第一枪，现在这两人后来居上，大有把自己甩下不管的意思。尤其是玻恩，仗着自己年长几岁，又有教授地位，经常勾引一些小青年来合作，可谓资源占尽，风光无限。历史已经加快了脚步朝前赶路，英国狄拉克在海森堡回德国的两个星期内就看到了玻恩和约当的“二人文章”，觉得这次这个文章写得比较数学，没有海森堡的“一人文章”那种乱枪打鸟的感觉，狄拉克很快就看懂了，既然已经看懂，不由得自己也有点手痒——他想自己写点东西。前面已经说过，狄拉克读大学的时候，出身于电机工程系，他是一个电机工程师，这种工程师背景的人生经历，为他的理论物理研究，提供了不可多得的洞见。盖因为电路系统，无论多复杂的网络，总可以看成是一个黑箱，它有一个信号输入端，还有一个信号输出端。电机工程师们习惯计算这个黑箱的传输函数。

按照维纳的控制论，人其实也是机器。因此如果把人看成一个电路系统，道理也是一样的，当有信号输入的时候，这个人一定会做出反应。比如这个人在街角莫名其妙被人暴打了一顿，他必然有所反应。但问题在于，如何计算出他会采取何种反应（计算响应曲线）——是扑上去咬打他的人一口，还是站在原地不动，或者飞身起来踹对方一脚，这不是很容易预测的。电机工程师们一直在处理这样的问题，那就是，当一个电路系统有一个信号输入的时候，它会输出什么？

为了处理这个类型问题，狄拉克在不久就引进了所谓狄拉克函数，

这个函数是一个冲击函数，所谓冲击函数就是这个冲击是在瞬间发生的，持续时间很短，但强度极大，它就好像闪电一样一出现就消失，那么这种冲击函数在傅里叶变换后具有平坦的功率谱——系统的传输函数就是输入函数为狄拉克函数时所对应的输出函数。狄拉克发现，这个函数不但在电路有用，甚至可以用在新生的量子力学里。

(2)

很多学物理的年轻人，内心深处都有一个问题：物理学的高潮过去了没有？

答案其实很简单，物理学的唯一一次大高潮早在1925年的时候已经来了。——和股票市场的牛市一样，物理学的高潮一般也只持续两到三年的时间，转身就进入熊市。不过想学习物理的学生，自然应该在熊市里入市。那时候物理学的高潮已经彻底来了。但大家都还没有完全准备好。

海森堡对玻恩说：“老师，我看了你和约当的二人文章了，要不我们三个人再写一篇综述文章，把量子力学的整个架构给建立起来？”

玻恩说：“好啊，我想我们已经完整地建立起了一个新理论。是应该整理出一个综述来。”

于是三个人马不停蹄地写出来了“三人文章”——《论量子力学（2）》，这种综述文章一般由这个领域的大牛人来执笔，基本上不是报告单个研究成果，而是写出最近的一系列研究前沿，给同行当教材来看的。



狄拉克等图书馆开门 绘画：张京

在这个文章中，矩阵力学大部分内容，被表达了出来，比如用厄米矩阵表示可观察物理量，微扰方法等，这些都用矩阵作为基本的数学工具，因此可以说当时的哥廷根大学物理系已经开始沉浸在一股用矩阵做成的迷雾之中。

在这个文章中出现的不对易关系 $pq - qp = (h/2\pi i)I$ ，玻恩他们引进了一个记号，写成了一个看上去很陌生的数学记号

$$[p \quad q] = I$$

狄拉克这时候人在剑桥心也在哥廷根，他也很快写了一篇文章，指出 $[p \quad q] = I$ 这个含有方括号的数学记号似曾相识，它其实和经典力学的泊松括号完全是类似的——不过当时的狄拉克不太记得泊松括号的精确定义，在度过了难熬的一个周末晚上，等星期一图书馆开门的时候才去查到相关资料，确证了这个记号确实与泊松括号非常类似。

如果说狄拉克比别人天才在什么地方，那就是他可以一眼就看出这泊松括号才是海森堡思想体系的精髓。他轻易地透过海森堡的表格，把握住了这种代数的实质——不遵守交换律。这种不遵守交换律的数学对

象可不仅仅在矩阵里出现过，他想起了什么？狄拉克的脑海里闪过一个名词，他以前在上某一门动力学课的时候，似乎听说过一种运算，同样不符合乘法交换律。但因为时间有点长远了，他还不是十分确定，他甚至连那种运算的定义都给忘了。那天是星期天，所有的图书馆都关门了，这让狄拉克急得像热锅上的蚂蚁。第二天一早，图书馆刚刚开门，他就冲了去找书看，果然，那正是他所要的东西：它的名字叫做“泊松括号”。

泊松1798年入巴黎综合工科学学校深造。在毕业时，因优秀的研究论文而被指定为讲师。受到数学家拉普拉斯与拉格朗日的赏识。他于1806年接替傅里叶出任巴黎综合工科学学校教授。1812年当选为巴黎科学院院士，他是法国第一流的分析学家，年仅18岁就发表了一篇关于有限差分的论文，受到了勒让德的好评，他一生成果累累，发表论文300多篇，对数学和物理学都作出了杰出贡献，泊松一生从事数学研究和教学，他的主要工作是将数学应用于力学和物理学中。他第一个用冲量分量形式写分析力学，使用后称为泊松括号的运算符号来简化一些偏微分计算记法。他所著《力学教程》在很长时期内被作为标准教科书。

总之，狄拉克发现，不必费九牛二虎之力去搬弄一个晦涩的矩阵，以此来显示量子力学体系和经典体系的决裂，实际上完全可以从经典的泊松括号出发，建立一种新的代数。这种代数同样不符合乘法交换律，狄拉克把它称作“q数”（q表示“奇异”或者“量子”）。所有物理量如动量、位置、能量等概念（甚至更激进的，时间与空间等基本物理量），现在都要改造成这种q数。而原来那些老体系里的符合交换率的变量，狄拉克把它们称作“c数”（c代表“可对易的”）。

简单地说，狄拉克发现量子力学与经典力学通过泊松括号相互挂钩，可以实行一揽子的学习计划。

“看。”狄拉克自言自语道，“海森堡的矩阵当然是对的，但我们不用他那种牵强附会高深莫测的方式，也能够得出同样的结果。用我的方

式，同样能得出 $xy - yx$ 的差值，只不过把那个让人看了生厌的矩阵换成我们的经典泊松括号罢了。然后把它用于经典力学的哈密顿函数，可以顺理成章地导出能量守恒条件和玻尔的频率条件。重要的是，这清楚地表明了，我们的新力学和经典力学是一脉相承的，是旧体系的一个扩展。 c 数和 q 数，可以以清楚的方式建立起联系来。”

狄拉克把关于泊松括号的论文寄给海森堡，海森堡热情地赞扬了他的成就，不过带给狄拉克一个糟糕的消息：他的结果已经在德国由波恩和约当作出了，是通过矩阵的方式得到的。

狄拉克并不为此感到很郁闷，显然他的法子更简洁明晰而且与故去的经典力学形式上一致，所以他觉得这样更美。

这个发现确实是美的，狄拉克如梦初醒，原来量子力学的不对易关系在经典力学里正是泊松括号的类似物。

这意味着什么呢？

经典力学和量子力学的关系的秘密就埋藏在这个类比中吗？

经典力学的运动方程， $\{ \}$ 表示泊松括号， H 表示哈密顿量，则力学量随着时间的演化方程可以写为

$$dA/dt = [A, H]$$

受到这个方程的启发，狄拉克发现，量子力学中力学量 A 演化满足的运动方程也满足相同的形式：

$$dA/dt = [A, H]$$

其中 A 和 H 都是算子（有限矩阵或者微分算子）。

能量矩阵用 H 来表示，称为哈密顿算子。
这就是矩阵力学中矩阵的运动方程。

外一篇 波函数么正演化

(1)

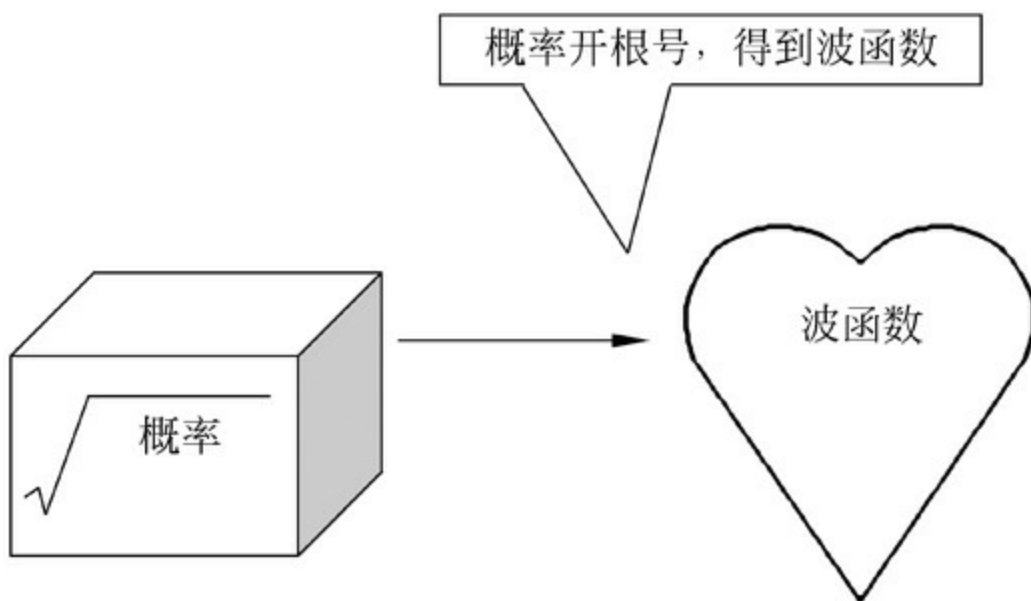
我们已经讲了矩阵的运动方程，那么我们还需要再回顾一下波函数的运动方程，这样才可以大致完整地勾勒一个新物理学的剪影。数学家们一开始不太理解1925年的新生的矩阵力学，倒不是说这个力学的数学很难，而是这个力学的物理解释很是费解。到了1926年，波动力学也出现了，薛定谔在做出波动力学以后，自己上去证明了波动力学和一年前的矩阵力学是等价的——他是在能量表象下做出这个证明的。既然矩阵力学和波动力学被证明是同气连枝。为了统一语言进一步理顺头绪，物理学家把矩阵力学和波动力学统一称为“量子力学”。

量子力学背后有一些人神共愤的新东西，这个潘多拉魔盒里出来了一个名叫“概率”的幽灵。概率的观念深入到理论的骨髓，大家似乎全看到了，上帝是以概率的方式在处理世上的一切。

“上帝难以捉摸”是爱因斯坦对量子力学的观感——这里的“上帝”，为斯宾诺莎的上帝，而不是基督教心中的那个耶和华。在量子力学中，一个原子的衰变是概率事件，一个原子的衰变方式也可能有多种选择，就好像一个苹果，可以分裂变成3个桃子，或者直接变成一个金苹果，所以，根本不存在确定性——这让爱因斯坦很不爽。

概率是一个数学概念，比如，天气预报会说，明天下雨的概率是40%，后天下雨的概率是30%。精确的概率定义则可以参考苏联的数学家柯尔莫哥洛夫的著作——总之，概率论是很严格的数学理论。

目前的中学数学课程经过改革，也是讲授概率的。量子力学，是一门更深层次意义上的概率课程，只不过，这个课程处理问题的方式在于先把概率开模平方得到波函数，然后找到波函数演化方程。



(2)

虽然概率是不确定的，但波函数演化的过程却是确定性的——物理学家称为“么正的”——么正性使得波函数在演化过程中保持概率守恒。

我们再把薛定谔的方程写出来 $i\hbar \frac{d\psi}{dt} = H\psi$

如果把H看成一个与时间无关的函数（虽然它其实是一个算子，但形式解答总是可以的），那么，上面这个方程是一个很简单的微分方程，这个方程具有如下的解答。

$$\psi(t) = \exp(-iHt)\psi(0)$$

很明显， $\exp(-iHt)$ 可以看成是对初始时刻波函数 $\psi(0)$ 的一个操作，这个操作被数学家称为“单参酉群”，也就是物理学家所谓的“么正演化”（酉和么正是同一个英文的翻译，么正2个字连读发音就是酉）。

么正？对外行来说，这个词语貌似很吓人。不必受惊，其实，简单地说，么正的意思是说，把一个矢量转动一个角度而不改变矢量的长度。打个恰当的比喻，这其实就像在酒吧里的转动吧台边的一把转椅那

样简单——只要屁股就可以决定其转动方向和角度，够简单。

波函数的演化，遵循的是薛定谔方程。演化是“幺正的”，也就是说，在演化过程中，波函数好像就是一把转椅，演化过程只不过是这把转椅转动了一定的角度，而不破坏这把转椅本身的结构。

薛定谔的波函数的幺正演化具有理解上的困难，如果电子确实如方程所描述的那样以波函数（比如说高斯分布）的方式弥散于全空间，那么薛定谔如何能解释电子枪打在电视屏幕上出现的一个个斑点？

换句话说，在薛定谔方程中，电子是一个波函数（比如说高斯分布），在各个地点都有分布，而在人们观测的时候，这个波函数瞬间就坍塌成为一个只在局部非零的狄拉克函数（delta函数）——仅仅出现在一个地方，而在其他地方为零。

（3）

我们来做一个比喻，比如说在北京，无论在三里屯，还是在什刹海，酒吧里的每一个转椅都是一个波函数——根据量子力学，任何有质量的物体都是波函数，当你转动转椅的时候，相当于在做幺正演化，可是，事情并没有那么简单。

假设这个时候，什刹海上暖风吹得游人醉，有一个绅士，春心荡漾的他在酒吧抱着一个暧昧的女孩子一起坐在转椅上，转呀转，觉得非常开心……

这个时候，演化依然是幺正的……

蓦地，从外面进来另外一个女子，这个人的瞳仁里有燃烧的火焰，这个人（其实是该绅士的妻子）狠狠地看了转椅一眼，绅士抱着小情人的转椅突然就崩溃了，转椅倒在地上——转椅坍塌了——波函数坍塌了！！！！

波函数是会坍塌的。波函数的坍塌，起源于观测者的观测。这在量

子力学中被称为“波函数的非么正演化”。“波函数的非么正演化”说明，量子力学理论虽然看上去很优美，与时间无关，与地点无关——但与人有关。

“波函数的非么正演化”和“波函数的么正演化”一起，才构成了整个波函数的演化理论。波函数已经是量子力学物理学家的灵魂所在，海森堡后来也深深地浸染了这种世界观，他为自己设计了一个墓碑，上面这样写着：“he lies somewhere here”，直译过来就是“他躺在这里，且在别处”。海森堡的意思是说，他死后，没有人观测他的时候，他的灵魂弥散在全空间，如果有人去观测他，则他就躺在这坟墓中。

13 科莫湖畔的费米

(1)

在海森堡发表了他的不确定性原理，狄拉克发表了矩阵的运动方程以后，他们都奠定了自己在物理学历史上不可动摇的地位。在物理学历史上还很少有其他物理原理可以与海森堡的不确定性原理相媲美。其他物理理论，比如爱因斯坦的相对论，也仅仅是描述物理世界的一级近似——正如牛顿力学里抛物线的运动轨迹往往是忽视了摩擦力作用的结果，在宇宙极早期的高温环境中，爱因斯坦的引力理论也不得不被修正。而海森堡的原理，非常确切地说明，物理世界其实是模糊的，远远没有大家想像的那么精确。

1927年的夏天，当海森堡去莱顿大学当教授，到莱顿大学以后，他马上成为理论物理系的系主任，到了9月，他要去意大利的科莫参加一次学术会议。这一次学术会议的对外宣传是为了纪念发明了电池的伏打逝世100周年。但其实海森堡相信，这次会议跟电池没有一毛钱的关系，组织者是为了搞到会议经费而搞了一个如此冠冕堂皇的理由。

意大利，我来了，海森堡心里暗忖，意大利物理在伽利略以后，已经死亡。

科莫是意大利北部阿尔卑斯山山谷中的一座小城，这个山谷里还有一片迤迤的湖光山色，这个湖就是科莫湖，类似于西湖，欲把西湖比西子，淡妆浓抹总相宜，这句诗同样适用于科莫湖，只不过科莫湖不在繁华的城市中，因此颇有一分野趣。

1927年金秋的科莫湖，一条游船在湖面上荡漾，上面有几个中年人和几个青年人。甲板上有三个小年轻跟在一堆大人后面也在湖上饱览这明媚的秋光。这三个人因为年纪很相仿，于是在一起很投缘，他们还拍

照留念了，从合影里可以看出，左边的分头是费米，中间的黄毛是海森堡，右边的方脸正是泡利。



费米、海森堡、泡利

“你们觉得玻尔这次在会议上讲的东西是不是开启了物理学的新时代？”费米小声地问。

费米是意大利人，和泡利同年纪，毕业于比萨师范大学（该学校虽然当时号称和巴黎师范齐名，都是拿破仑设立的，但实际上却是一个二流本科学校）。几年前他曾经去过哥廷根，在那里他觉得自己远远比不上海森堡等人，所以，回到意大利以后总觉得自己低人一等——他也深刻感觉，意大利的物理水平，确实已经腐朽。

“玻尔这次讲的互补原理，太哲学了，但是我喜欢。”泡利说，“不过这样的原理我一天至少能写出三个来。”

“也许是这样的，玻尔的互补原理，其实就是我半年前发现的不确定性原理，——当时，他去挪威滑雪了，回来搞了一个互补原理，说什么无论在辐射还是物质中，波动性和粒子性是相互排斥又是相辅相成的——说实话，这个原理确实似乎想把薛定谔的那套波动力学和我们的矩

阵力学统一起来，”海森堡冷静委婉地说，“.....虽然我不喜欢薛定谔——这次他没有来，他要是敢来，我就要猛烈吐槽波动力学——不过，我还是相信，物理学真的正在发生革命.....原子世界也许真的需要波动与粒子在一起的描述。”

费米听闻海森堡那不太确定的语言，一阵糊涂一阵明白，有点窃喜。因为海森堡的感觉似乎和自己完全一样，接下来问：“海森堡，你觉得如果爱因斯坦参加这次会议，他听了玻尔的互补原理，会有什么反应？”

海森堡沉默了一下，说：“爱因斯坦是大师，他的想法一向诡异，我们不好简单判断。不过，也许爱因斯坦并不喜欢玻尔的这一套哲学——这套哲学说实话就是中国的那套阴阳鱼哲学——也就是所谓太极理论。”

费米大骇，问道：“阴阳鱼？”

旁边的泡利插话说：“阴阳鱼哲学，在东方很是流行，就是一个圆里有两条鱼，一条是白色的，一条是黑色的，但是，白鱼的眼睛是黑色的，黑鱼的眼睛是白色的。这个阴阳鱼哲学认为，世界上所有事情，都是白中有黑，黑中有白。”

(2)

费米从科莫回到罗马以后，心情有点低落，因为在科莫会议上，玻尔和玻恩等人根本就无视他的存在——这是一种无声的藐视，费米觉得自己在意大利这样腐朽的物理学环境要混出个人样来还真不容易。玻尔和玻恩等人，其实是看不起他的，费米还是隐约觉得自己的物理不行——比起同龄人泡利和海森堡等人，自己似乎总缺少什么。虽然几年前他得到了关于自旋为 $1/2$ 的粒子的统计学。可是，这还不足以引起大师们对他的兴趣。

我们当代意大利青年做物理真的不行吗？自从父亲和母亲去世以后，费米的家里也变得空荡荡的（十多年前，费米的哥哥夭折了），在这个世界上，现在他唯一的一个亲人就是他姐姐。

费米想去自杀。

“祖国啊，我的死是你害的。”

不过，这个念头一晃而过，因为他知道自己已经是意大利最牛的青年物理学家了，自己要是自杀了，还有谁能独当一面呢？

于是，他坚持着继续工作，为了生活多一些乐趣，他买了一辆法国产的波日奥小汽车用来带女朋友出去郊游——这辆车花了他大部分积蓄。

这个时候，总是有理的墨索里尼已经上台的，这个充满理想的中年人也正在试图改变意大利。墨索里尼梦想着建立一个新罗马帝国。政府用金钱鼓励人们结婚生子。男孩子从童年开始，国家法西斯党就一直向他们灌输着意大利的命运以及战争的美妙。每一个男孩子都必须佩带刺刀，到20世纪30年代中期，墨索里尼说：“意大利已经拥有800万把刺刀。”

年轻的费米教授的波日奥小汽车是蛋黄色的，非常好看，是一种敞篷车，最高时速30公里，拐弯的时候不能减速，开在路上绝对像一个鸭蛋黄在横冲直撞。

这辆车就停在罗马大学的校园里，顿时吸引了很多女学生的眼球。

罗马大学19岁的名叫劳拉的漂亮女生，是一名海军将军的女儿，这是一个白富美，犹太血统，以前鬼使神差地和费米一起爬过山，算是隐约认识费米，她也被这辆小车给吸引住了，心想：“这家伙做司机比做丈夫更适合。”

于是她经常参加费米他们的聚会，在聚会上费米经常和塞格雷等人在一起谈论量子理论。劳拉自然是听不懂，但她内心深处有一些欢喜。

费米正在慢慢地架构他的罗马学派，聚会喝酒打闹只是一种表面形

式，其实，学派的灵魂是新生的量子力学，他最杰出的徒弟是塞格雷，也是一个物理的狂热爱好者。

费米一开始也经常开车拉着他的罗马学派的几个哥们儿，加上劳拉和她的姐姐安娜一起出去郊游。几次以后，费米只拉着劳拉一个人出去兜风了，车子在山上颠簸，两人情投意合。

车停在山麓边，周围连一个人影也没有。

费米说：“这里好安静啊。”

劳拉羞涩地说：“是啊。”

费米小声地说：“我喜欢你。”

劳拉听见自己的心撞如小鹿突奔，说：“什么？”

费米说：“啊？没有什么？”

劳拉说：“哦，我也喜欢你……”

费米把手从方向盘上拿下来，放在劳拉的胳膊上，过了一会儿，他的手开始慢慢地游离……

回到罗马城里，爱情瓜熟蒂落。

劳拉的姐姐安娜，是一个画家，对数学物理很害怕，她每次看见费米，总跟她妹妹说：“瞧，你的对数又来了。”

劳拉娇嗔着说：“姐，别瞎说。”

1928年7月19日，27岁的费米结婚了，新娘是劳拉。

(3)

劳拉的出现，罗马学派的年轻人非常高兴，因为这样费米就可以安心地做学问了。

1927年，费米和玻尔一样心里颇不宁静，他们同时在思考以前的一个问题，13年前，查德威克就发现，有的原子会放出一种电子流，这个电子流来历不明，并且电子流的能量是非常不确定的，能量可以从零一

直到一个最大值。这就是历史上著名的“ β 射线连续谱”问题。

新婚之夜。

这个夏天的夜晚，格外地闷热，费米显得并不高兴。

在床上，妻子问他：“你似乎有心事？”

费米说：“我在考虑 β 射线连续谱问题。”

妻子说：“这问题那么重要吗？”

费米说：“是的，谱为什么是连续的呢？”

妻子说：“存在即合理，就是连续的。”

费米愣了一下，说：“为什么是连续的呢？”

妻子说：“我不知道，我不管——我妈给我们买的家具怎么样？”

费米说：“很好，可惜家具的腿怎么都是弯的？”

妻子说：“弯的比直的好看。”

费米说：“也许吧。你说，真奇怪， β 射线的能谱为什么是连续的呢？”

妻子说：“明天我们去度蜜月吧。你教教我物理学。”

蜜月如期而至。费米带着劳拉在罗马机场坐上了水上飞机，飞机贴着海面飞行，去往避暑胜地热那亚。

飞机沿着海岸线飞行，极目骋怀，海岸的沙滩上是彩色的阳伞和穿着比基尼的美女。

“教授，你该教我什么是量子力学了。”妻子说。

“好啊，我先给你讲讲电磁学吧。其实，麦克斯韦发现，光是电磁波……”坐在边上的费米侧过脸来说。

“我不信！”妻子说。

……

飞机像一只海鸥，掠过人们的头顶。大家都抬头仰望，欢呼雀跃……

在热那亚换乘火车，费米带着新婚燕尔的妻子在阿尔卑斯山南麓的

山间穿行。远处的雪山洁白，在湛蓝的天穹下闪闪发光。

他们找了一间乡村旅店住下。

这对幸福的旅人又开始了愉快的交谈。

“光是怎么产生的？”妻子问。

“一个电子，从2层楼跳楼了，跌在地上，就产生了光子。”费米开始变得通俗起来。

“啊，这样啊，那电子为什么要跳楼呢？”妻子问。

“你这个学生，真是不好教啊。”费米笑着说，“这叫量子跃迁，至于为什么发生跳楼（跃迁），这也许是因为电子受到了一些莫名的影响。物理学家把这个影响叫做微小的扰动。”

“哦，我相信。”妻子接着问，“那么，中子也会跳楼吗？”

费米说：“中子？你的意思是不带电的粒子吧？我猜想它也会跃迁的，它跳下来，就会变成带正电的质子，然后放出带负电的电子，然后……”

妻子趴在费米身上，贴着费米的背说：“我不信……”

物理学家面临困境，在中子转变为质子放出电子的过程（ β 衰变）中，能量似乎并不守恒。

结过婚的女人是很清楚的，家里有多少钱。如果老公把家里的钱偷偷拿给外面的“小三”花，那么钱的数目就会显得很守恒。

三体运动可以解释这一切，于是，泡利认为，在 β 衰变的过程中，放出的粒子中，除了质子和电子，还有一个看不见的“小三”在花家里的钱，这个在暗处的“小三”被称为“中微子”。在泡利这个解释之前，玻尔也有一个解释，玻尔认为，在微观过程中，比如 β 衰变中，能量守恒定理是失效的，换句话说，家里的钱本来就不守恒——爱因斯坦听说这件事情后，心想，你们都没有吃过狗屎，我爱因斯坦吃过，那就是我看到过玻尔的能量不守恒定理。爱因斯坦写信给别人说：“如果玻尔是对的，能量不守恒，那我宁愿成为一个鞋匠。”

但是泡利引进的中微子根本就没有被观测到过！！！！

这个偷偷花家里钱的“小三”是很会捉迷藏的，她人藏起来了，但钱照花不误。并且她花钱的数目是非常随机的，今天花100，明天花200，后天花500，大后天花50，根本没有一个特征谱，也是一条很光滑的连续谱！！！！

泡利描述 β 衰变的方程如下：

$$n = p + e + \nu$$

n是中子，p是质子，e是电子， ν 是电子中微子。

根据量子力学的角动量合成规律，很显然，这四个粒子都是自旋为1/2的费米子。并且电子和中微子不可能一开始就呆在原子核里面。

（用海森堡不确定性原理就可以计算，原子核那么小的尺度里不可能有自由电子，否则电子的能量会非常大。）

费米对这个小三问题很感兴趣，因为 β 衰变电子的连续谱曲线必须得到完整的解释。

他后来给出了这四个费米子直接相互作用模型，这是一个非常直观的有效理论模型，但完美地解释了 β 衰变电子的连续谱。这是薛定谔解出氢原子能谱以后量子理论最伟大的胜利，虽然这个费米的成名之作并不太完善，因为这个理论只在能量较低的时候才是对的，当能量较高的时候，费米模型算出来的散射截面大得不得了，与实验并不符合。但要等到后来人们才知道，这个小三问题所涉及的 β 衰变属于弱相互作用，四个费米子直接相互作用的这个理论是不可重整化的，按照物理学家们的说法因为旋量场的量纲是3/2，4个旋量场的量纲为6，所以费米耦合常数的量纲是-2，量纲为负数的量子理论，都是不可重整化的（所谓不可重整化，意思就是不存在可以抵消那些无穷大的项）。我们可以举一个例子来说明重整化，从北京到杭州，你可以坐飞机过去，大概需要3个小时，就可以到西湖边上的岳王庙。这里面有两个过程，在飞机

上，你花了2个小时从北京首都机场到了杭州萧山机场，走了900公里。在萧山机场下了飞机以后，到西湖边的岳王庙，你乘坐出租车，大概走了50公里，却花了1小时——但是，这就是说，飞机和出租车实际上是对距离的两种不同尺度的度量，还存在其他各种不同的度量，而可重整化的意思是说，你采取一定的交通工具，你可以到达西湖边的岳王庙。如果理论是不可重整化的，就好像一只蚂蚁从北京要爬到杭州西湖边的岳王庙，因为蚂蚁所要经过的路途对蚂蚁来说实在是太遥远了，细节太复杂了，要过黄河长江，要穿过山脉丘陵，要走过田野和池塘，这些复杂的微观地理使得蚂蚁根本不可能从北京爬到西湖边的岳王庙，因此，这个对于蚂蚁来说，从北京跑到杭州的路程实际上是无穷远的，这就相当于是不可重整的。

四个费米子直接相互作用的这个理论是不可重整化，但这个理论是解释中子衰变成质子和电子以及电子中微子的弱相互作用的有效理论。作为一种直接的相互作用模型，后来才被引进了传递弱相互作用的W和Z玻色子的理论所取代，这些W和Z玻色子的质量在80~100GeV量级，所以物理学家认为，发生弱相互作用的特征能量是GeV（这就好像说人类社会的特征尺度是一米，因为人的身高大约在1米附近，而刘慈欣的科幻小说《三体》的尺度是以光年为单位的，比如三体人所在的半人马星座离开太阳系的距离是4光年，因此在光年尺度上我们谈论火车与飞机的速度谁快是没有意义的。）GeV也就是10亿个电子伏特的能量——等于用一亿伏特的高压加速一个电子所得到的动能。但是，可以通过海森堡的测不准原理估计出，在原子核里面的环境温度，大概是在MeV这个能级上，这样的环境温度实际上大约等于200多倍电子质量（这个问题也可以由介子质量估算），这个环境温度差不多只有中子或者质子质量的1/10，在这样的环境下，因为温度太低，从量子力学概率的角度来说，弱相互作用发生的概率就很小，所以称之为弱，这个时候经常发生的另外一个相互作用是强相互作用，发生的概率是1，所以原子核

基本被强相互作用束缚在一起不会发生爆炸。

总之，年轻时代的费米所建立的直接的相互作用模型就好像物物交换的原始社会，到了现代社会，出现了货币，货物交换的模型就是间接相互作用模型，是比较合理的。但直接物物交换，也是有效的。

费米在这时候奠定了他在新罗马帝国的教皇地位，这个地位不可动摇，因为他深入到了原子的核心，而且这从此成为他一辈子的主旋律，一直到他在美国芝加哥大学领导的实验小组实现了人工控制的核爆炸，他的名字终被全世界所知道。

14 第五届索尔维会议

(1)

制碱的人，多数都会发财。

中国有一个制碱的大师，就是福建人侯德榜。在欧洲一位比利时的实业家欧内斯特·索尔维也是富可敌国。当然，后来侯德榜破解了索尔维制碱的工艺秘密，不过那时候索尔维早已经赚了很多钱，索尔维赚了钱以后，就资助了一个高规格物理学年会，这就是索尔维会议。每三年一次，相当于物理学界的武林大会，要广发英雄帖，号召各大门派的物理学领袖来参加。



第一届索尔维会议历史照片

到了1927年，已经是第五届索尔维会议，费米还有点默默无闻，不在被邀请之列，于是他缺席了。参加会议的部分人员名单如下：

1. 德拜。1884年出生于荷兰。1901年进入德国亚琛工业大学学习电气工程，1905年获电子工程师学位，因他通过偶极矩研究及X射线衍

射研究对分子结构学科所作贡献而于1936年获诺贝尔化学奖金，他曾经在索莫菲处学习，算是泡利等人的师兄。

2. 布喇格（W.H.bragg，1862—1942）。现代固体物理学的奠基人之一，他早年在剑桥三一学院学习数学，曾任利兹大学、伦敦大学教授，1940年出任皇家学会会长。由于在使用X射线衍射研究晶体原子和分子结构方面所作出的开创性贡献，他与儿子W.L.布喇格分享1915年诺贝尔物理学奖。父子两代同获一个诺贝尔奖，这在历史上恐怕是绝无仅有的。

3. 爱因斯坦。20世纪最伟大的科学家，被公认为人类历史上最具有创造性才智的人物之一。他的名字与相对论密不可分，其实，相对论包括两种理论：其一是他1905年提出的狭义相对论；其二是他1915年提出的广义相对论。

4. 埃伦费斯特（P.Ehrenfest，1880—1933）。荷兰物理学家，他的两个学生提出了自旋的概念，他本人是爱因斯坦的铁哥们。

5. 狄拉克（Paul Adrien Maurice Dirac，1902—1984）。英国物理学家，后来用数学方法描述电子运动规律时，发现电子的电荷可以是负电荷、也可以是正电荷。狄拉克猜想，在自然界中可能存在一种“反常的”带正电荷的电子。

6. 薛定谔（Erwin Schrodinger，1887—1961）。奥地利理论物理学家，与爱因斯坦、玻尔、玻恩、海森堡等一起于20世纪20年代后期，发展了量子力学。因建立描述电子和其他亚原子粒子的运动的波动方程，获得1933年诺贝尔物理学奖。

7. 康普顿（A.H.Compton 1892—1962）。在1922—1923年间研究了X射线经金属或石墨等物质散射后的光谱，发现X射线的波长在散射以后发生改变。

8. 泡利（Wolfgang E.Pauli，1900—1958）。奥地利出身的犀利的物理学家，海森堡的师兄，其父亲是维也纳大学的物理化学教授，是著

名的泡利不相容原理的提出者。

9. 海森堡（Werner Karl Heisenberg, 1907—1976）。德国理论物理学家，量子力学第一种有效形式（矩阵力学）的创建者。

10. 玻恩（M. Max Born, 1882—1970）德国理论物理学家，量子力学的奠基人之一，爱因斯坦的朋友，与玻尔在同一个量子力学矩阵形式战壕里的老战友。

11. 玻尔（bohr, niels）。1885年10月7日生于丹麦首都哥本哈根，父亲是哥本哈根大学的生理学教授。从小受到良好的家庭教育，1903年进入哥本哈根大学学习物理，1909年获科学硕士学位，1911年获博士学位。他是量子力学领域的一代宗师，也曾是学校足球队选手。

12. 普朗克（Max Planck, 1858—1947）。近代伟大的德国物理学家，量子论的奠基人。普朗克其实是一个很悲惨的科学家，他大儿子死于“一战”，2个女儿死于难产，小儿子被希特勒枪毙。一家人中他是最后死的，多次白发人送黑发人。

13. 居里夫人（1867—1934）。波兰出身的女物理学家。她曾两次获诺贝尔奖，1903年的物理学奖，1911年的化学奖。1893年获物理学位，1894年获数学学位，1903年获博士学位。居里夫人以放射性作为论文题目，她研究了很多物质，发现钍及其化合物的特性与铀相同。研究沥青铀矿时，她发现了镭和钋。1910年她成功地分离了纯镭。

14. 洛伦兹（Hendrik Antoon Lorentz, 1853—1928）与塞曼（Pietr Zeeman, 1865—1943）。因研究磁场对辐射现象的影响、发现塞曼效应，分享了1902年度诺贝尔物理学奖。

15. 朗之万。1872年1月23日生于巴黎，法国著名的物理学家，德布罗意的老师，爱因斯坦相对论理论“双胞胎悖论”的提出者。

在会议之前，江湖上已经风传哥本哈根学派已经给出了量子力学试卷的一套标准答案，包括三部分：

1. 玻尔互补原理

2. 海森堡的不确定性原理

3. 玻恩对波函数的概率解释。

对于这套纲领性文件，其他门派表示不怎么赞同，肃杀的氛围掺杂着假繁荣的基调，让人心神时刻都不得安宁。1911年，第一届索尔维会议在布鲁塞尔召开，以后每3年举行一届。1927年，第五届索尔维会议在比利时布鲁塞尔召开了，洛伦兹是此次物理学家峰会的主持人——他早年在莱顿大学任教期间创立了电子论，并与塞曼因研究磁场对辐射现象的影响，发现塞曼效应，分享了1902年度诺贝尔物理学奖。1904年他提出狭义相对论中著名的洛伦兹变换公式，并指出光速是物体相对于以太运动速度的极限，而且他比爱因斯坦等人年长20岁，是公认的德高望重之人。这次会议的爱因斯坦与玻尔两人就量子力学进行了一次大辩论。

(2)

这次会议的正式安排是从1927年10月24日到29日，为期6天。会议的主题是“电子和光子”。会议议程如下：首先劳伦斯与布拉格作关于X射线的实验报告，然后康普顿报告康普顿实验以及其与经典电磁理论的不一致。接下来，德布罗意作量子新力学的演讲，主要是关于对应于有质量粒子的德布罗意物质波。随后由波恩和海森堡分别介绍量子力学的矩阵理论，接着由薛定谔介绍他的波动力学。最后，则是量子力学宗师玻尔在科莫演讲的基础上再次做一个关于量子公设和原子新理论的报告，进一步总结互补原理，给量子论打下整个哲学基础。这个议程本身就是量子力学发展史，从中可以明显地分成三派：只关心实验结果的实验派：布拉格和康普顿；哥本哈根派：玻尔、波恩和海森堡；还有波动派系：德布罗意、薛定谔、爱因斯坦。

会议在热烈与友好的气氛中开幕，按照议程，大家先就康普顿的实

验做了探讨，这里面其实也有中国物理学家吴有训跟康普顿做的工作，当然当时在会议上是没有人提到吴有训这个中国人的名字的。在这个关于X射线衍射的讨论以后，气氛开始变得对立，零星的战火开始烧到德布罗意那里，因为他试图把粒子融合到波的图像里去，提出了一种“导波”（pivot wave）的理论，认为粒子是波动方程的一个奇点，它必须受波的控制和引导。这个理论被泡利狠狠地批评，德布罗意招架不住。

薛定谔以夸父追日的方式做物理，已经有很多年，对他来说，女人比诗歌重要，诗歌比物理重要，所以薛定谔永远无法在失落情怀中走出来，在这次会议上，薛定谔做了一个叫“波动力学”的报告，也算是一个他个人对量子力学的总结。

在讲台上，这个诗人开始了他的演说：“尊敬的洛伦兹教授，尊敬的爱因斯坦教授，各位物理学家，教授们，上午好。未曾开口，已觉空虚.....在演讲之前，我先要谈谈对海森堡和玻恩的理论的一点看法，我觉得我其实并不认为量子力学已经是一个完善的理论，也不太理解玻恩给我的波函数做的所谓概率解释.....我这个波动呀，其实不是真实空间里的波动.....你们也许听说了.....前年，海森堡去哥廷根数学系演讲，希尔伯特他老人家没有听懂，后来希尔伯特曾经问过他的助手冯·诺依曼，矩阵力学到底是什么玩意，——现在据说他们已经有了研究结论，这个波函数是生活在希尔伯特空间里的.....”

台下的人听到这里，连爱因斯坦也吃了一惊。

薛定谔好像是一个摆好了擂台的擂主，口若悬河，说话的声音大起来了：“各位，今天我们所建立的量子力学理论，也许是一个悲剧.....”

台下开始骚动起来，泡利小声地对旁边的海森堡说：“这个淫魔在说什么？”

海森堡侧身过来，耳语道：“他是一个诗人，不过我听他的话，怎么那么别扭！他似乎想朝我们伟大的哥本哈根学派开炮。”

“.....各位，我的波函数其实就好像是一朵云彩，但它带有电荷，

至于电荷是不是均匀分布，我不知道，但是，我想我愿意承认，波函数就好像天边的云彩……”诗人薛定谔说，“电子就好像是这一朵云彩，非常地漂亮……”

众人听得耳朵里嗡嗡响，云彩……云彩……

(3)

就在各人精彩的报告开始变得意兴阑珊之时，索尔维会议结束了，结束的那天下午，洛伦兹走出会堂的楼梯，步履有点蹒跚。他已经很老了，老得记不清楚很多事情，他甚至想不清楚一件事情，那就是狭义相对论到底是谁发现的，是他，还是伏瓦基（Voigt），抑或是爱因斯坦的老婆，或者可能是菲兹杰拉德？难道是庞加莱？

在众人走出会堂到广场上合影的时候，他看到爱因斯坦似乎如众星捧月一般被人围观，他明显感觉到这中年人身上散发出来的成熟稳重的气息，而自己居然已经老朽成这般模样。

按照江湖规矩，散会了要合影留念。座位已经摆好了，第一排有9个座位，谁坐在中间呢？所有的人内心都在打鼓。这是非常有讲究的。因为坐在最中间的这个人，地位必须足够高，而这必须要得到大部分与会者的默认。这种事情虽然大家不会明显地讲出来，但当一个人真正坐在中间第5个位置上的时候，他一定能明显感觉到一种君临天下的威仪。

近了，渐渐地近了……

洛伦兹终于开口了，他先讪笑了几声，然后亲切地说：“爱因斯坦，你坐中间吧。”

爱因斯坦犹豫了一下，说：“好吧。”

……

等大家坐定，露出将要笑着离开的神态，摄影师喊了一句：“好！”

请各位老师保持笑容！”

闪光灯一闪！



等大家各自回到大学，合影照片已经洗出来了，海森堡是一个很敏感的人，他发现，爱因斯坦坐在前排最中间的位置上。洛伦兹则在左边第四个位置上。第一排中，还有居里夫人和普朗克等人，依次排坐在爱因斯坦的两翼——这看上去确实是自人类诞生以来的最强阵容。

换句话说上一代的武林盟主洛伦兹已经把位置让给了爱因斯坦。

洛伦兹24岁的时候，那是在遥远的1878年，他在母校当大学老师，那时候爱因斯坦还没有出生呢。

洛伦兹一开始是做电磁学的，他希望自己能从微观的角度把宏观的麦克斯韦方程推出来。

显然，在麦克斯韦的方程中，电场和磁场全是宏观量（几何量），从微观的角度来看，是一个平均场的效果，因为电场一般是由一些微观的点电荷激发的，那么从微观角度来说，有点电荷的地方其实有很多电场的奇性——这就好像早晨上班高峰地铁站里的人流，从宏观上看来，人流是比较均匀连续流体，但对于地铁站里拥挤着的美女来说，人流并不那么均匀——总有一些小伙子会故意贴着她的身体。

洛伦兹费了九牛二虎之力，终于证明了，如果假设电荷有微小的粒子附带，那么，麦克斯韦方程确实可以从微观角度做一个平均场给推出来。

洛伦兹把那些带有电荷的微小的粒子，称为电子。

洛伦兹的《电子论》在汤姆孙发现电子之前就出现了。《电子论》有一种思想倾向认为电荷是由微小的粒子附带的。这种想法很像纯净水一样天然无味。但人是不能超越时代的，如果当时的洛伦兹能够在数学上证明电荷是量子化的，那他才可能具有超越爱因斯坦的地位。

但洛伦兹还做了其他事情。

1880年开始，就有一个叫麦克耳孙的美国海军军官，是一个硕士，在欧洲他搞了一个实验，企图证明地球绕太阳系的公转速度会影响到他设计的干涉仪的干涉条纹。

可惜，麦克耳孙的实验总是得到零结果，也就是说，光的传播速度并不会和地球的公转速度简单叠加上去——这其实就是狭义相对论的全部意义所在：“光速在任何参考系都不变。”麦克耳孙的实验也引起了洛伦兹的注意。洛伦兹证明，当把麦克斯韦的电磁场方程组用伽利略变换从一个参考系变换到另一个参考系时，真空中的光速将不是一个不变的量，从而导致对不同惯性系的观察者来说，麦克斯韦方程及各种电磁效应可能是不同的。为了解决这个问题，洛伦兹提出了另一种变换公式，即洛伦兹变换，用洛伦兹变换，将使麦克斯韦方程从一个惯性系变换到另一个惯性系时保持不变。当时写出洛伦兹变换的还另有他人，各自有自己的思路，比如伏瓦基就是想要波动方程的算子在某个变换下形式不变来研究问题的，经过痛苦挣扎，他也写出了洛伦兹变换。

无论怎么样，历史的发展如浪奔浪流，洛伦兹变换是两个参考系之间的线性变换，这个变换开始在江湖上流传起来，但其物理解释，他们都没有解释对。

洛伦兹的研究工作，总离开真正靠谱的解释差之毫厘，但也足以谬

以千里——比如他对塞曼效应的解释是经典谐振子上加上一个电磁力引起的经典频率的交错，并且因为这个不太靠谱的解释得到了诺贝尔奖金。他把自己得到的洛伦兹变换看成是一个绝对静止参考系和一个相对匀速运动参考系之间的变换——可是，宇宙中有绝对静止的参考系吗？这自然是没的。

于是，洛伦兹在糊涂地做学问的过程中渐渐老去。

人是往地底下走的，第二年，洛伦兹就死了。德高望重的洛伦兹离开了滚滚红尘。

(4)

洛伦兹的离开，留下很多东西。

他影响了一个数学家的成长。

事情是这样的。

1910年，wolfskehl基金悬赏的费马大猜想的钱还没有奖出去，于是，这个基金产生的利息就用来请科学家到哥廷根去做演讲。洛伦兹作为电子论的数学大师，相对论坐标变换的提出者，自然有这个机缘出现在哥廷根大学的讲堂之上。

“今天，我开始讲本次系列讲座的第四讲，物理学中的新问题和旧问题是我们的主题……”洛伦兹在台上说。

台下有一个年轻人，正竖起了耳朵听，突然，他听到洛伦兹说：“……基于金斯的黑体辐射理论，人们已经知道的一点是，黑体辐射的光是电磁波，于是，一个很自然的推论就在1905年由英国的金斯推出——这是一个驻波条件，任何吉他手都是很清楚的——吉他高手必须要改变手指按琴弦的位置，才能改变乐音基频。吉他基频对应的波长 λ 的半整数倍等于弦长 L 。



绘画：张京

同样道理，按照这个经典图像，在一个密闭容器（炼钢炉）中，电磁波的所有模式中，反弹形成驻波的模式才是基本的，能量在这些模式之中平均分配——这就是金斯的错误的黑体辐射谱。我们考虑波动的驻波模式，那么在这个模型中，很显然的是，电磁波的基本频率模式是与容器的外形相关的，零频时被称为调和问题，一般模式是亥姆霍兹方程，总之，问题与边界的形状有关系……在座的各位数学家，我想，这个问题对你们来说，也许很简单，……这个问题就是解答振动方程，我们能不能通过鼓的声音，来反推出鼓面的形状？……”

台下听讲座的年轻人，正是外尔，他是希尔伯特的学生，功夫了得，他马上知道，这个“听音辨鼓”的反问题很有意思。

他再也坐不住了，洛伦兹还在台上讲自己对这个问题的一些物理上的猜测，但外尔觉得自己应该马上退出江湖，关门去搞这个反问题。他突然感觉自己得了躁郁症，站起来，朝门外走去……他要马上动手了。

一年后，外尔很好地解决了这个问题的大部分，他得到的结果是：

$$\lim_{\lambda \rightarrow \infty} n(\lambda)/\lambda \sim S$$

λ 是鼓面振动发出的声波的基本波长（特征波长）， $n(\lambda)$ 是比 λ 小的特征波长的个数， S 是鼓面面积。

虽然这个数学问题在形式上远离了黑体辐射的理论，但源头却是金斯的黑体辐射驻波条件。这也是量子力学理论对数学的反作用，推动了数学的发展。

外尔有自己的想法，他珍藏洛伦兹带给他的物理遗产，在1928年写了一本书，这书的第一版是在哥廷根用德语出版的，影响力不强，第二版是在普林斯顿改写的，被翻译为英文，讲的是群论在量子力学中的应用，这本书，自然为物理学家们理清楚了头绪。哥廷根数学学派给最初的量子力学提供了数学解释。

15 狄拉克矩阵：相对论与量子力学的婚姻

(1)

量子力学的问题不但与数学关系密切，也与当时物理学的另外一个新兴学科相对论之间可以建立一个桥梁，这个桥梁是由狄拉克担任总工程师施工架设的。

狄拉克对物理学的主要贡献是发展了量子力学，提出了著名的狄拉克方程，并且从理论上预言了正电子的存在。狄拉克原来从事相对论动力学的边缘研究，自从1925年海森堡访问剑桥大学以后，狄拉克深受影响，把精力转向量子力学的研究。1928年他把相对论引进了量子力学，建立了相对论形式的薛定谔方程，也就是著名的狄拉克方程。这一方程具有两个特点：一是满足相对论的所有要求，适用于接近光速运动的电子——在原子内，电子的运动速度大约是光速的百分之一这个数量级，所以不需要考虑狭义相对论效应，因为狭义相对论效应只在接近光速的情况下才体现出来；二是它能自动地导出电子有自旋的结论，也就是说，自旋是一种相对论效应。这一方程的解很特别，既包括正能态，也包括负能态。狄拉克由此做出了存在正电子的预言，认为正电子是电子的一个镜像，它们具有严格相同的质量，但是电荷符号相反。狄拉克根据这个图像，还预料存在着一个电子和一个正电子互相湮灭放出光子的过程；相反，这个过程的逆过程，就是一个光子湮灭产生出一个电子和一个正电子的过程也是可能存在的。1932年，美国物理学家安德森在研究宇宙射线簇射中高能电子径迹的时候，奇怪地发现强磁场中有一半电子向一个方向偏转，另一半向相反的方向偏转，经过仔细辨认，这就是狄拉克预言的正电子——最近2013年的物理学大新闻就是华人物理学家丁肇中领导的实验小组发现了冷暗物质衰变为正负电子的迹象。在1932

年以后物理学家很快又发现了 γ 射线产生电子对，正、负电子碰撞“湮灭”成光子等现象，全面印证了狄拉克预言的正确性。狄拉克的工作，开创了反粒子和反物质的理论和实验研究——以至于很多日本动画片里都有所谓的反物质武器或者狄拉克海的概念。他的主要著作有1930年出版的《量子力学原理》。

狄拉克早年在布里斯托尔大学读电机系的时候，狄拉克觉得自己简直是进入了一个野鸡大学，那里的学生毕业了以后就是当电机工程师，所以多数人缺乏深邃的思考，这让狄拉克陷入孤独的境地。

他的那些同学们总研究输入和输出问题，把一切电路系统看成一个黑箱子，甚至还把女孩子看成是一个黑箱子，问如果输入的是精子，输出的是孩子，那么，子宫作为一个黑箱的传输函数到底是什么？他们还问其他的问题，比如女人分成两个种类，纯情和骚情，那为什么同一个女人总是既纯情又骚情？

狄拉克心想，再没有比布里斯托尔大学更猥琐的大学了。

不过关于传输函数和女人种类的问题，深刻地改变了狄拉克。他意识到作为输入函数， δ 函数是可以定出传输函数的，而至于同一个女人为什么有不同的侧面，他感觉这似乎是一个所谓表象的理论。不过，在当时，关于这个问题的细节，他还没有想清楚。

1921年，狄拉克和几个工科大学生一起聆听了关于相对论的一系列讲座，激励他更深入地思考时间和空间的关系。他起先埋头钻研爱丁顿1920年出版的畅销书《空间、时间和引力》（Space, Time and Gravitation）。在随后的布里斯托尔大学数学研究结束之前，他已掌握了狭义相对论和广义相对论，包括里面所使用的绝大多数数学工具。

大学毕业以后，他找不到工作，于是，就进入了剑桥大学物理系读研究生。

到了剑桥，那里的学生素质就真不一样，大家总是讨论学术问题——比如四色问题或者完美正方形问题，反正大家是把数学当作乐趣来

钻研的。

狄拉克心想，“我喜欢剑桥，这里很好。”

数学系的哈代教授还记得有一次他去医院探望那个印度数学天才拉马努扬，拉马努扬生于印度Erode，他十岁之时才首次接触数学，但表露出过人的数学才华。拉马努扬无心学习其他科目，以致考试经常不及格，加上家境贫困亦令他无法专心学习。拉马努扬结婚后利用他的计算才能寻找工作，最后在会计师楼上班。他希望完全专心于数学，1912—1913年间拉马努扬写信给数位剑桥大学的学者，而只有当时在三一学院的哈代发现拉马努扬的天才。当时哈代亦研究数论，且已是此范畴中世界首屈一指的学者，但哈代依然对拉马努扬的研究成果深表佩服。哈代最后邀请拉马努扬到英国。对于拉马努扬的公式，他认为“一望而知，只有最出色的数学家才能写下这些公式。它们一定是真的，因为根本没有人可以有如此的想像力去无中生有。”哈代曾提到他本人对数学最大的贡献，就是“发现了拉马努扬”。他认为拉马努扬的天才可以和欧拉等巨匠相比。拉马努扬最后在剑桥大学三一学院获得教席，亦成为英国皇家学会的会员。但是因为拉马努扬自小已是体弱多病，他又醉心研究，始终无法适应英国的气候及水土，最后得了包括肺结核病在内的多种病症。最后他不得不于1919年回到印度，不久便与世长辞，终年三十三岁。而流传在剑桥的故事是这样的，有一天哈代对拉马努扬说他坐的出租车号牌是1729，他说“这数字真没趣，希望不是不祥之兆”。拉马努扬却说：“不，1729是一个相当有趣的数字，它可以写成两对不同立方数之和，而在拥有这特性的数字中，1729是最小的一个。”这个故事在剑桥早已成为美谈，当狄拉克进来的时候，他也听说了这个有点装X的故事。

到了1928年，狄拉克已经发展了量子力学表象理论，表象理论的核心思想是把量子力学的波函数看成是一个抽象的矢量，生活在一个抽象的希尔伯特空间之中。换句话说，他已经理解到大学时代那些同学们在讨论的那些关于女人的话题——女人其实是一个希尔伯特空间里的抽象矢量，你要想了解这个女人，可以把她投影到商场，也可以把她投影到厨房，反正，在不同的地方，她会有不同的表现。离开具体语言环境谈论女人，是毫无意义的。

这已经是登峰造极之作了，但狄拉克还是深深地为两件事情苦恼：

1. 完美正方形（拉格朗日四平方和定理的高级版本，哈代肯定也在思考这个问题）。存在不存在一个以整数为边长的正方形，它的面积可以被分裂为4个小的不相等的整数边长的正方形之和？

2. 薛定谔的方程不是洛伦兹不变的（不满足自己的偶像爱因斯坦提出的狭义相对论）。存在不存在一个方法，把波动方程的算子开根号，得到一个一次方的算子？

一切尽在想像。一个晚上，他梦见自己找到了解决第2个问题的方法：其实，要想对相对论性的波动方程的算子开根号也许是可以实现的，你可以假装已经开了根号了，……

对于费米来说，通往上帝铺设的林荫道去接近上帝的最短路线是通过泡利。当泡利搞出自旋为 $1/2$ 的粒子时，费米发展了它们的统计学，当泡利提出中微子假设的时候，费米也马上做出 β 衰变的4费米子理论。狄拉克也是如此，他时刻关注着泡利的动作，因为在他的潜意识中，泡利是上帝派驻在人间的代表。

我们再回顾一下当时的情景，泡利从塞曼效应的复杂的光谱数据中得到了一个很大的猜想，这个猜想简直可以与那个既不能被证实又不能被推翻的哥德巴赫猜想相提并论。泡利猜想说：“任意两个自旋为 $1/2$ 的粒子，不能占据同一个量子态。”这就是泡利不相容原理——到1940年才被泡利自己证明出来（利用所谓自旋统计关系，这来源于量子场论能

量正定加上洛伦兹不变性以及哈密顿算子的厄米性）。当在没有被证明之前，泡利的猜想无往而不胜，江湖上已经把它鉴定为真理。

泡利得到了非相对论性自旋的表示——也就是著名的泡利矩阵。因为本书是科普读物，读者群默认为高中生为主，所以，在这里要缓慢地解释一下泡利矩阵。在物理学里，一般把矢量 \mathbf{M} 写成3个分量的线性组合。

$$\mathbf{M} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + a_3 \mathbf{e}_3$$

其中， \mathbf{e}_1 ， \mathbf{e}_2 ， \mathbf{e}_3 是基矢量。

如果要求， \mathbf{M} 自己和自己的内积如下（初中生的数学）：

$$\mathbf{M} \cdot \mathbf{M} = a_1^2 + a_2^2 + a_3^2$$

则相当于要求，基矢量满足如下条件：

$$e_i^2 = 1$$

$$\{e_i, e_j\} = 0$$

这里 i 和 j 取1，2，3。 $\{ \}$ 表示正对称符号。

这样的基矢量，自然是可以用矩阵表示出来的。对于 1×1 的矩阵表示，就是大家熟悉的直角坐标系的3个基矢量。对于 2×2 的矩阵表示，就是泡利找到的3个矩阵。

泡利是用三个 2×2 的矩阵来表示非相对论自旋 $1/2$ 的电子的。

这是一个伟大的重复发明，因为数学家早已经在搞这个 $\mathfrak{su}(2)$ 李代数了—— $\mathfrak{su}(2)$ 李代数当时给人的感觉好像阑尾，数学家并不清楚其功能。泡利出现以后，代表上帝赋予其意义。泡利矩阵正是 $\mathfrak{su}(2)$ 李代数。

狄拉克自然很清楚泡利的工作。泡利得到自旋 $1/2$ 的表示，却没有说明为什么要有自旋。这让狄拉克心花怒放，他觉得应该马上上去搞一把。很快，他发现，满足狭义相对论的克莱茵高登波动方程自然是洛伦

兹不变的，那么开根号以后，依然是满足狭义相对论的，只不过这个时候时间和空间的地位是平等的，指标也就变成了4个。但他会遇到泡利遇见过的性质类似的问题，那就是要求找到一个代数的表示。

$$\{e_i, e_j\} = \pm 1$$

其中i和j取1, 2, 3, 4。

“泡利当年为了得到泡利矩阵，是做矢量的平方，现在，我反过来，是开算子的根号……当时他在三维欧几里得空间里做，现在我在四维平坦时空里做，除此之外，我与他之间还有什么区别呢？……”狄拉克在暗中淌泪，“难道我真的不能超越泡利吗？难道我这辈子，仅仅是为了把泡利矩阵改名为狄拉克矩阵吗？”

无论怎么样，先把这个狄拉克矩阵（相对论性自旋的表示，自动包含反粒子）写出来吧。最简单的写法是4×4的四个矩阵。写出来以后，就得到了狄拉克方程，“方程比人还要聪明”——狄拉克发现，这个方程里，不但有自旋，而且还有负能量的电子。

(3)

当时，狄拉克就傻了，怎么办？开根号出负数是初中生都知道的东西，但现在出现负的能量，怎么解释？如果电子可以朝负能量跃迁，而负能量又没有最小数值，那么所有的电子都可能跃迁到负能量。这样的话，宇宙是不稳定的——换句话说，如果股票市场不存在一个市场底，那么所有股票的价格可能要跌到零，甚至跌成负数。这样的股票市场与屠宰场还有什么区别？再说了，如果存在负的股票价格，那么买进股票的人不但花了钱而且还欠了上市公司一屁股的债，那买股票的人，是真正的傻呀。不可能，一定不可能，必须存在一个市场底。

狄拉克心想，我要引进一个猜想，类似于泡利猜想。于是，狄拉克也提出了一个猜想，这个猜想是这样说的：“负能量的电子海已经被充

满。”

在英伦这个岛国的四周，到处是满得快溢出的海水，海鸥像和平鸽一样在天边盘旋，其实这个地球上到处都是山，只不过有些山谷被水填满，也就成了海。

外一篇 量子力学中的相位

狄拉克作为量子力学的集大成者，其深邃也迥异于常人。后来有人问狄拉克，在量子力学中，是测不准原理最重要吗？狄拉克回答说，不是，在量子力学中相位才是最重要的。

这到底是什么意思呢？

如果你只知道虚数单位 $i = \sqrt{-1}$ （当然也可以是一*i*），那么你是否觉察不出下面这个式子的错误？

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{(-1)} \cdot \sqrt{(-1)} = i^2 = -1$$

如果你在一念之差选择了放弃学习物理或者是数学，是否你一辈子也不会知道问题的答案？如果多年后一个阴郁的下午，你在大学的课堂上摆弄着手机，你完全忘记了这是一门叫做“数学物理方法”或者是“复变函数”的课程，这时候讲台上白发苍苍的老教授在黑板上重新写下这个你今天第一次看到的等式，你是否会记得你今天对物理和数学的兴趣？

话扯远了，其实上面这个等式只不过是复变函数中的一个简单的习题。它当然是错误的，错误就在于 $\sqrt{-1}$ 的值不止一个，而是等于

$e^{i\left(k + \frac{1}{2}\right)\pi}$ ， $k=0, 1, 2, 3\dots$ ，它有无限多个值。这，就是复数的世界！

因为有了离散的群体，所以我们有了整数；因为整数的比值有可能

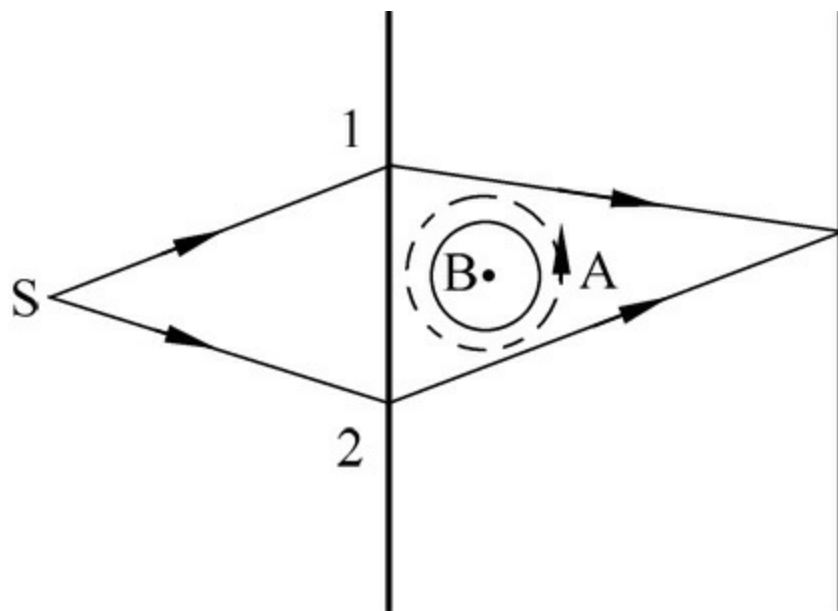
不是整数，所以有了分数；因为边长为1的正方形对角线的长度既不是整数也不是分数，于是我们有无限不循环小数——他们组合在一起，就成了实数；如果负数也可以做幂运算，我们就可以把实数再扩大为复数。毕达哥拉斯学派的西帕索斯因为发现了无理数而遭到残害，卡当引入了复数获得荣誉，世事就是这么无常，同样是扩大了数域，在不同的时代就有不同的命运。

一个复数由实部（实数部分）和虚部（虚数部分）构成，比如 $z = x + iy$ ，其中 x, y 是实数。利用著名的欧拉公式 $e^{i\theta} = \cos\theta + i\sin\theta$ 也可以把复数写成 $z = \rho e^{i\theta}$ 的形式，其中 $\rho = \sqrt{x^2 + y^2}$ 是 z 的模，

$\theta = \arctan \frac{y}{x}$ 称为 z 的幅角。这有点像是把直角坐标变换到极坐标的感觉，在相同的 ρ 不同 θ 的情况下“矢量”的长度并没有改变。于是物理学家管这个 $e^{i\theta}$ 称为相位因子，起初还看不到它有多大的能耐，但是随着时间的推移科学的进步，它的作用越来越大起来。就连著名物理学家、诺贝尔物理学奖获得者杨振宁先生也说过：“.....量子化、对称和相位因子是20世纪物理学的主旋律。”那么，这个连“矢量模长”都改变不了的相位，究竟是怎样在物理学中产生影响的呢？

在量子力学中一个微观粒子体系的状态，用一个波函数 ψ 来完全描述，而波函数 ψ 是希尔伯特空间中的一个矢量。在空间中找到某粒子的概率是 ψ 模的平方 $|\psi|^2$ ，也就是概率只跟这个矢量的长度有关，而它的相位，在取模平方的过程中将被消去。而实验中能测得的物理量只是波函数模的平方，也就是概率。所以在量子力学的早期，有些物理学家们并没有重视这个相位因子。直到一个实验现象浮出水面.....

1959年，阿哈罗夫和波姆发表了一篇论文，其中提到电子干涉中，波函数 ψ 的相位可能会受到磁矢势的影响。第二年就有人做了这个实验，他就是钱伯斯，一个实验物理学家。那么，什么是AB效应呢？



电子衍射的装置

上图中S是电子源，1、2是挡板上两个极小的狭缝，狭缝之后有承接屏。承接屏上有荧光物质，当电子打在其上时会发出光亮，就可以知道电子主要落在屏上的什么地方。而B是一根通电螺线管，磁场被封锁在管内。实验先在没有B的情况下进行电子干涉，在承接屏上得到电子干涉图样。然后加入B，B是一个小范围的、非常靠近挡板的螺线管，所以不会阻挡到电子的运动。从理论上来说电子并没有经过磁场，承接屏的图样应该不会有什么变化。但是实验事实却让人惊讶，承接屏的干涉图样改变了，电子的运动发生了变化！

这个实验最爆炸性的结论是仅用磁感应强度描述磁场不是完备的，磁矢势也会有物理实在。但是我们从另一个角度去想想，为什么电子打在荧光屏上各处的亮度改变了？你可以说是打在某处的电子数量减少了，而打在另一处的电子数量增加了。但是考虑到电子的波粒二象性，我们应该认为这是相位在作祟。既然电子是波，那么就会发生相干叠加增强和减弱；实际上这个电子的干涉实验就是在证明这一点。而在波的叠加中，并不是简单的振幅相加，而是必须考虑到相位的相干叠加。形象地说，这种叠加不是 $1+1=2$ ，而是 $x \leq 1+1 \leq y$ 这样一种形式。具体等

于多少，这要看所处位的相位。所以AB效应造出的影响，是改变了电子的相位，使得在某处本来发生相长干涉的地方发生了相消干涉，造出承接屏图样的变化。

此后对于量子力学相位的研究越来越多，取得的成果也越来越多。特别是1984年英国物理学家贝利发现在几何相中的一种依赖于含时参量的相位，称为贝利相，更是在物理学的其他领域如量子信息、凝聚态、量子光学等有着重大的应用。而关于量子力学的相位的研究，现在依然是前沿的问题。因为，正是因为量子力学的相位，才有了量子力学的相干叠加，才有了不同于平常世界 $1+1=2$ 的量子世界的千奇百怪。

16 海鸥，马约拉纳

(1)

每一只海鸥都是一个死去水手的灵魂。

假如你去一个海滨城市，在浪漫暧昧的夜晚，徐徐海风吹拂之下，一些游轮餐厅会放着这样的音乐，那是Rod Stewart的SAILING：

I am sailing, I am sailing...cross the sea.

I am sailing stormy waters, ...

...

Can you hear me, can you hear me, through the dark night far away?

I am dying, forever crying, to be with you; who can say?

...

Oh Lord, to be near you, to be free.

20世纪20年代中期，费米还在意大利刚开始经营他的新罗马帝国，有一个极端帅气清秀的年轻人马约拉纳是他的学生。当时他俩都只有二十几岁，年轻而有梦想。

马约拉纳说：“每隔500年才有一个类似阿基米德或牛顿这样的科学家出现，而每隔100年就会有1至2个爱因斯坦和玻尔这样的人出现。”

费米说：“那费米是几百年出一个？”

马约拉纳说：“我们谈的是爱因斯坦与玻尔……”

马约拉纳21岁的时候加入了罗马大学物理研究所由费米领导的研究组，这个组里汇集了一批意大利当时最优秀的青年物理学家。大家紧密团结在费米周围进行工作，唯有马约拉纳是一个单打独斗的人。不过他

的超级分析与计算能力及天才的物理直觉对整个研究组的帮助是无法估量的。马约拉纳确实是个计算天才：只要有他在的地方，就没有人会用计算尺和笔算。只要需要问他，请告诉我1538的对数，或者，243的平方根乘以578的立方根等于多少，他都能很快得到答案.....费米和他曾有过一场计算能力的PK，费米用纸笔计算尺，马约拉纳只用脑子，然后比赛结果是平局.....而众所周知，费米的计算能力也是非常强的：后来在美国爆炸原子弹的时候，他就站在很远处，手里拿张纸，撕成碎纸片。原子弹的冲击波来了，他把碎纸一扔，然后根据纸片被卷走的高度、速度和距离计算释放的能量值。冲击波走了，他就算出来了，而且计算出来的结果和精密仪器测试的结果不相上下.....有一次费米正在计算中子的寿命，在黑板上算啊算，马约拉纳就站在旁边想啊想。最后费米说我得到结果了，结果他听到马约拉纳说，“算出来是十五分钟吧。”费米当时就郁闷了，觉得马约拉纳计算能力远比自己强，他把粉笔一甩，就走出去了。

马约拉纳对自己的研究工作自我要求很高，他很少发表研究成果，除非他本人觉得研究结论已经无可挑剔。1929年，费米研究组的成员之一瑞萨缇（Franco Rasetti）在美国加州理工学院进行访问研究时，测量了氮原子核的自旋数，其结果却是“1”。马约拉纳立刻意识到原子核应该是由带正电的质子和一种不带电的、质量与质子相近且自旋亦为 $\frac{1}{2}$ 的粒子所组成。他把这种粒子叫做“中性质子”，也就是人们后来发现的中子。具体到氮原子核，其组成应为七个质子加七个“中性质子”。他同时也意识到，为了使原子核不致因内部的质子们“同性相斥”而分崩离析，核内一定存在一种比电磁力强得多的相互作用力，他称为“交换作用力”——这就是后来人们所说的强相互作用力。他建构的这套原子核稳定性理论可以说是现代量子色动力学的前身。可是不知出于何种原因，尽管费米费尽唇舌，马约拉纳却始终拒绝把这个理论拿出去发表。

几个月后，俄国的物理学家伊万年科（Dmitri Ivanenko）认识到了中子的存在，海森堡也发表了与马约拉纳非常接近的理论。费米抱怨他坐失良机，马约拉纳也仅仅一笑置之——从这里也许可以看出马约拉纳是一个比较淡漠的青年，也许还有点抑郁。

马约拉纳还是第一个想到宇称可能不守恒的人——这个问题就是后来李政道与杨振宁得诺贝尔物理学奖的研究课题。

宇称是一种描述粒子在空间反演变换下性质的物理量。就像我们有的人习惯用右手写字，有的人习惯用左手一样，基本粒子的自旋也具有类似的特性——左旋与右旋。对大多数粒子来说左旋与右旋是对称的，即如果存在具有左旋的A粒子，就一定也存在具有右旋的A粒子。假如一个具有左旋的A粒子去照镜子，镜子里看到的就是一个具有右旋的A粒子。在粒子的相互作用中，如果以左旋粒子取代同种的右旋粒子而结果不变，这种相互作用就具有左与右的对称性。粗略地说，这就是宇称守恒。

宇称只在弱相互作用过程中会不守恒，而弱相互作用一般都涉及中微子。中微子有非常奇特的性质——都是左旋的。也就是说，如果一个左旋中微子去照镜子，镜子里则什么都没有，因为右旋中微子根本就不存在。

马约拉纳是研究中微子的理论家，他当时就已经极有可能已经意识到了中微子的这种高度不对称性。他在20世纪30年代初开始建构的中微子理论。到了1941年中国的物理学家王淦昌也开始研究中微子，并且提出测量方案。

1932年，由于健康原因，马约拉纳辞去了在罗马大学物理研究所的职位，开始了长达四年几乎足不出屋的“闭关”生活。在这四年里他没发表过什么论文，却完成了一批杂七杂八的小型研究，包括地球物理、电子工程、数学和相对论。到了1937年，在没有任何征兆的情况下，他突然“破关”而出，去应聘意大利巴勒莫大学的一个教授职位。在应聘要求

当中有一条是必须提交一篇论文，于是他就将那篇已经尘封五年之久的有关中微子的文章拿了出来，这样才使世人有机会一睹他的极富想象力的中微子理论。

1938年3月25日他给家人和他任职的那不勒斯大学物理研究所所长卡瑞利（Antonio Carrelli）各留了一封短信后，就登上了一艘开往西西里首府巴勒莫的邮船。一般人和警方都把这两封信解读为绝命书。不过也有两件事令人费解——他支领了半年的薪水并带走了所有重要的科研笔记，这不大像一个准备自杀之人所为。尽管如此，如果事情到此为止，人们大都还是会认定他自杀了。可出人意料的是，他平安抵达了巴勒莫，而且又发了一封电报和一封信给卡瑞利。电报仅一句话“别紧张，信随后就到”，信里则明确说他放弃了自杀的念头。根据记录，他确实买了返回那不勒斯的船票，而且有个同舱人（三人住一间舱房）曾作证说，他在那不勒斯下船时，马约拉纳还在舱里睡觉。但马约拉纳却从人间蒸发了，没人确切知道他是否在那不勒斯下了船，甚至连他到底上没上开往那不勒斯的船也是个未知数。这种不确定的结局，为后人留下了想象的空间。以致几十年来不断有人宣称在世界的不同角落遇见过马约拉纳，版本之一是：在20世纪60年代初，他经常在智利的一个小酒馆里吃饭，还在餐巾纸上演算数学问题……这些传闻无一得到证实，恐怕均是媒体的炒作。时至今日，意大利人也没有忘记他。他多次成为科幻小说或电影的主角，甚至还有关于他的科幻连环画集。在连环画里，马约拉纳的结局最为辉煌——被外星人接走了！

总之，很多年以后，马约拉纳跳上帆船出去航海，意大利的海也是那么蓝，渐行渐远，……马约拉纳消失在海平面之下……物理学并不是摇奖机，但物理学家的命运似乎也有很多面，没有人知道马约拉纳是自杀了，还是失踪了？——江湖上有流言说，马约拉纳决定消失是因为他已经预见到原子弹将毁灭地球，他只想把音容笑貌留在人们的记忆里——而假如真是这样的，那么原子弹的制造者费米显然有不可推卸的责

任，这真是“兄弟一场，我不杀伯仁，伯仁因我而死”。

(2)

狄拉克在1928年得到的四分量旋量是一个复函数，四个狄拉克矩阵也是复的。狄拉克猜想，负能量电子海已经充满了，那么如果负能量电子海里的电子被激发为正能量电子，显然会挖出一个萝卜（负能量电子）留下一个坑（缺少一个负能量带负电荷的电子，相当于多了一个正能量带正电荷电子）——于是，狄拉克把这个萝卜坑理解为反粒子。

狄拉克做完这些工作以后，他在物理学上的贡献就已经定型。他给世人留下的印象就是高度自闭。所以当波兰青年物理学家英费尔德（此人后来成为爱因斯坦的助手）1933年前后来到了剑桥时，福勒就建议他跟从狄拉克研究正电子理论中的一个问题。英费尔德于是去拜访狄拉克，下面就是他后来记述下来的那次会见：

“我穿过圣约翰学院那狭窄的木楼梯，然后敲响了狄拉克的房门。他悄悄地把门打开，并以友好的手势指着一张带扶手的椅子。我坐下来并等着他讲话，但鸦雀无声。我告诉他说，我几乎不会讲英语，他友善地笑了笑，但没有作任何回答。”

英费尔德试图把谈话引向深入，且把福勒的建议告诉了他。

“仍然没有回答。我等待了一会儿，接着直接提出了问题：‘你反对我同你一道研究那个问题吗？’——‘不。’我终于让他吐出了一个字。接下来，我开始讨论那个问题，并拿出钢笔来以便写出一个公式。他一言不发地站起来，拿来了纸张。但是，我的钢笔却写不出字来，这时他又悄无声息地拿出他的铅笔，并递给了我。我再次问他一个直接的问题，而得到的回答是5个字，我花了两天的时间才消化它们。谈话即将结束，我试图延长它。‘当我遇到困难的时候，你不介意我来麻烦你吗？’——‘不。’我离开了他的房间，奇怪而又沮丧。他太难令人亲近

了，要是我事先就知道每个剑桥人都是这么认为的，那么我就不会有什么不愉快的感觉了。如果英国人都觉得他很特别，那么对一个喜欢嚼舌根的波兰人就更可想而知了。”

狄拉克当然不是无礼，尽管他的举止被英费尔德描述得好像是无礼的。他作为逻辑的耕耘者，只不过是按逻辑行事，因此在某些场合就表现得与社会所接受的方式背道而驰。他直接回答问题，但不作评论或者其他不得要领的陈述。他接下来会坦率行事，在一些场合下，这种坦率并不总是被认为是友好的表示。

研究四分量旋量的马约拉纳也继承了狄拉克的怪异风格。他似乎对意大利学术界除了费米以外的那些傻子们的集体无意识深恶痛绝，决定自己教育自己——马约拉纳有一句名言：“物理已入歧途，我们都已入歧途。”于是，他在思考怎么样才可以在江湖上扬名立万——当时，他脑子里有2个命题：

1. 存在不存在一个整数 Z ，这个整数的相反数是它本身。
2. 存在不存在一个整数 Z ，这个整数的倒数是它本身。

他发现，对于以上两个命题，整数 Z 都存在，答案分别是0和1。

马约拉纳于是决定把这个朴素的思想应用到新生的量子力学里去。

1937年马约拉纳告诉费米说：“我能找到四个实矩阵来表示狄拉克代数。我找的四分量旋量也是实数形式的。——换句话说，存在一个没有电荷的旋量粒子，它的反粒子是它本身。”

费米说：“听上去很不错，请去发表吧。”

于是马约拉纳就正式发表了他的一个猜想——物理学历史上，牛人是很喜欢做猜想的，前面也已经看到过了，比如泡利猜想，狄拉克猜想，一个好的猜想可以扼住命运的咽喉。马约拉纳的猜想也像一只蚊子一样能在午夜飞行，弄得很多年轻人晚上睡不着觉，只听见耳边嗡嗡作响：“自然界存在一种有质量无电荷的自旋为 $1/2$ 的粒子，它的反粒子就是它本身。”

其实早在1929年，数学家外尔就已经描述了一种带自旋但没有质量的粒子，被称为外尔旋量。而马约拉纳的猜想表面上看似乎像是一种数学游戏，马约拉纳所要求的这种旋量会在一些特定维度的时空中被找到——实际上这意味着人们面对的量子世界可以有更多出人意料的结果，比如超对称理论中，引力子的超对称伴侣就是一种自旋为 $3/2$ 的带质量的马约拉纳旋量。旋量的故事既然已经展开，也许不免让一些文科读者花容失色，但故事将越来越扑朔迷离。

简单地说：

1. 没有电荷的自旋粒子，用马约拉纳旋量描述。
2. 没有质量的自旋粒子，用外尔旋量描述。
3. 没有电荷没有质量的自旋粒子，用马约拉纳-外尔旋量描述（但在四维时空不存在）！

自旋，电荷，质量。这3个量子数就好像是自助餐厅里的3道菜，你可以自由选择吃哪几个！

马约拉纳的出现就好像一个海鸥，最后隐没在茫茫大海，但他的出现说明费米学派已经成为一个培养杰出人才的新学派了，费米不愿意自己成为一只海鸥，他要飞得更高。

17 朗道：苏联之子

(1)

和煦的夏风吹着哥本哈根的玻尔，他的两个儿子正在茁壮成长。看在眼里，玻尔真是高兴啊，当时他并不知道几年以后大儿子将夭折。



研究所的院子里宿草盈阡，爬山虎爬满了墙壁，绿油油的一片藤蔓之下，世界各地的年轻人来了又走，真是应了那句老话“铁打的营盘流水的兵”，虽然海森堡已经离开哥本哈根去了莱比锡大学，但玻尔相信，只要自己还存在，那么，哥本哈根就好像一个工厂，会制造出一个又一个海森堡。

没有错，1931年的研究所里，又来了几个年轻人，比如说，伽莫夫、特勒，还有朗道。

特勒（Teller）站在院子里，听见蝉的叫声，感到无边的寂寞，他发现自己还真不是一个做学问的料，那两个苏联人太牛了，自己简直像一个傻子一样混在哥本哈根。

那两个被特勒嫉妒的年轻人，正是伽莫夫和朗道。

1908年出生的朗道真是非常年轻（比那帮1901年出生的人小了7岁），他显然已经来迟了一步，量子力学的楼船早已经扬帆起程，船上尽是笙歌艳舞一片，歌妓已经被别人拥抱，留下朗道一个人在江边兴叹：“我来晚了一步，我妈怎么搞的。”他19岁的时候，还在苏联，就思考一个问题，因为海森堡说物理量是矩阵，而薛定谔说波函数有一个密度分布。朗道19岁做物理就好像29岁一样老练，他马上把矩阵和密度联系起来了，创作了一个伟大的新概念——密度矩阵。

这就是朗道，他自然是心比天高，他顺路到哥本哈根，看到自己的一个师兄伽莫夫也在这里。

朗道说：“师兄，你最近在研究什么？”

伽莫夫说：“隧道效应啊。对了，你叫什么名字？”

朗道心想，天下谁人不识我朗道，真是瞎了眼了，你居然不知道我的名字：“我叫朗道，今年23岁。去年我研究了电子在磁场中的运动，得到了朗道能级，你不知道吗？”

伽莫夫说：“哦，原来你就是朗道，太好了，朗道能级是什么？你给我讲讲！”

朗道说：“你居然不知道朗道能级！朗道能级啊，说来话就长了……”

（2）

18世纪末19世纪初，有一个叫安培的人，思考环行电流——电流像北京二环地铁那么流动。安培研究的结果很是惊人，环行的电流能产生一个类似于条形磁铁产生的磁场（磁偶极场）。在安培以后，数学家太喜欢环形电流产生的磁场了，于是高斯等人就思考，能不能把两个通电线圈相互套起来（像两个戒指那么相互套起来，或者两个手铐相互铐起

来），然后计算一下空间的磁场分布，看看能不能在磁场中读出电流线圈的拓扑结构——相当于说，看到一缕光线，你能不能推断出发光灯丝的形状。到了1931年，拓扑在物理学中已经开始变得重要。英国的狄拉克也在思考有没有磁单极场——存在这样一个磁场，通过观察这个磁场，你会发现，这个磁场是起源于一个点，而不是类似于安培的环线。

简单地说，23岁的朗道发现，关于安培的环路电流可以用量子力学来重新做一做。高中物理学里的洛伦兹力表明，在均匀磁场中，电子将作圆周运动。那么，如果把量子力学加进来，这个图像是什么样子的呢？

显然，电子的圆周运动可能会导致电子能量的离散化——周期性的运动总会导致量子的离散现象，所以女性的生理周期也会导致其脾气的多能级爆发——这个离散化的能量，被称为朗道能级。

朗道能级是很容易解出来的，因为圆周运动其实是两个简谐振动的合成，只不过现在的情况有了微妙的变化，在量子图像里，圆周运动的圆心坐标（a，b）非常特殊，a与b是不对易的。

$$[a, b] \neq 0$$

朗道把自己的发现过程跟伽莫夫讲了一遍，伽莫夫听到圆心坐标不对易的时候，一个头有两个大。虽然他相信这个不对易的原因是由于磁场的存在。但这个结果还是很令人吃惊的，因为磁场存在以后，几何学被彻底改变了，需要用非对易的坐标来描述一个点的位置，那么，点的x坐标和y坐标就是不可以同时确定的……

伽莫夫说：“师弟，也许你做错了什么。但我不知道你哪里错了。”

朗道说：“师兄，我回苏联以后，再考虑这个问题吧。这个问题是很有意思的，对了，你还要回苏联吗？”

伽莫夫说：“你先回去吧，梁园虽好，非久留之地，苏联需要你……”

1931年，伽莫夫被召回苏联，任命为列宁格勒科学院首席研究员，并在列宁格勒大学担任物理教授。当时斯大林制度下，伽莫夫感到自己富于想象力的天性受到压制，很不开心。1933年出席在比利时布鲁塞尔召开的一次会议时，伽莫夫抓住机会离开了苏联。离开苏联后，伽莫夫在法国巴黎的居里研究所从事研究，1934年移居美国，在密歇根大学担任讲师，同年秋天被聘为哥伦比亚特区华盛顿大学的教授。在华盛顿大学工作期间，伽莫夫主要从事宇宙学和天体物理学研究，发展了大爆炸宇宙模型，并且研究了宇宙初始阶段化学元素起源的问题，这个时期是他学术生涯的顶峰，取得了一系列重要的研究成果。

现代宇宙大爆炸理论是在1932年由比利时牧师勒梅特首次提出的。20世纪40年代，伽莫夫与他的两个学生——拉尔夫·阿尔菲和罗伯特·赫尔曼一道，将相对论引入宇宙学，提出了热大爆炸宇宙学模型。热大爆炸宇宙学模型认为，宇宙最初开始于高温高密的原始物质，温度超过几十亿度。随着宇宙膨胀，温度逐渐下降，形成了现在的星系等天体。他们还预言了宇宙微波背景辐射的存在。40年代，伽莫夫指派阿尔菲研究了大爆炸中元素合成的理论。在阿尔菲1948年提交的博士论文中，伽莫夫说服了汉斯·贝特把他的名字署在了论文上，又把自己的名字署在最后，这样，三个人名字的谐音恰好组成前三个希腊字母 α 、 β 、 γ 。于是这份标志宇宙大爆炸模型的论文以阿尔菲、贝特、伽莫夫三人的名义，在1948年4月1日愚人节那天发表，称为 $\alpha\beta\gamma$ 理论。

(3)

早在1929年，玻尔邀请爱因斯坦到哥本哈根作学术报告，朗道恰好也在此访问。爱因斯坦刚讲完，他便站起来指出其中一个错误。略作沉思之后，爱因斯坦这样认错：“大家可以把我今天讲的全部忘掉。”因为表现过于突出，列宁格勒大学在1934年免去朗道的答辩环节，直接授予

他理学博士和数学博士学位。后来，他受好友卡皮察之邀，到物理问题研究所主持理论物理方面的工作。个性傲慢再加上心直口快，最终给这个狂妄的年轻人带来灾难。1938年4月，一辆黑色轿车接走了朗道。因为随口表达对社会的不满，又被发现参与起草一份地下传单，朗道被直接送进安全部门，以“德国间谍”的罪名判处10年徒刑。为了营救他，玻尔专门给斯大林写信，恳求赦免，但并没奏效。

最终卡皮察押上自己的身家性命，向斯大林保证朗道不再从事任何反革命活动。更重要的是，他告诉这位最高领导人，自己在低温领域的研究获得重大进展，急需理论家帮助，而苏联只有朗道从事这方面的理论研究。或许是斯大林认识到了朗道的价值，就在朗道觉得自己“再在监狱待半年必然会死掉”时，他在1940年获准保释出狱。

从此，处处都有密探监视他的一言一行，并汇报给安全部门。这个自称“有学问的奴隶”的人，知道自己的言行会关系到好友卡皮察的安危，开始主动远离政治。每当身边有人谈起政治，他总是自我嘲讽道，“我是懦夫”。

“二战”后，卡皮察认识到原子弹的危害，拒绝参与核武器的研制，被赶出莫斯科。朗道则参与了研制，当别人躲着卡皮察时，只有他每月前去拜访。但“懦夫”还是管不住自己的嘴巴。密探在1947年呈送的报告里说，朗道抱怨国家的科学界彻底卖身求荣，“而且比国外是有过之而无不及，国外的学者毕竟还拥有某种自由”。他更不满的，是周围人的“卑鄙无耻”，“不仅科学家如此，评论家、文学家、报纸杂志记者也如此。只要有人给钱，上边让他们干什么他们就干什么”。这些报告并没有令朗道失去自由，倒是1953年斯大林的死讯，让朗道误以为他彻底获得了自由。他立即停下手头的核武器研究，并宣称“他死了，我不怕了，所以我不干了”。他不知道，新上台的赫鲁晓夫也不信任他，他依旧是现实中的囚徒。



朗道在狱中的照片

总之，朗道从哥本哈根回国以后，因被怀疑是德国间谍而入狱。

在监狱里，朗道感到非常愤怒，他怒发冲冠。

中午，放风的时候，他总是在高墙之下仰望井口那么大的天空，心里大骂斯大林。

狱卒把一盘饭放在他面前，吆喝道：“快吃吧，窝囊废。”

朗道气得直哆嗦，用手指着狱卒的鼻梁骂道：“畜生！我是朗道，等我出去，我要……”

朗道在监狱里发现了一个巨大的秘密，是大家都知道但心照不宣的。那就是电子在均匀磁场中，如何与磁场耦合（相互作用）的呢？磁场是一个经典场，电子是一个量子化的粒子，它们的相互作用到底怎么写呢？

虽然江湖上总有人在说什么最小耦合，但语气又有点像阿Q嘴巴里的“革命”，语意是模糊的。

自己还来得及，虽然量子力学的楼船已经出发，那些西方人在船上已经醉生梦死，乐不思蜀，但作为苏联之子的朗道，依然可以在岸边另开一片天地——实际上他研究电子和磁场，开创了凝聚态物理学的大片

新河山。可以说，在那个国度里，他其实一直在等待出狱，尽管他在物理学上的成就彪炳青史。

18 广岛之吻

(1)

从西方到东方，隔着一个苏联。

从西方的物理学到东方的物理学，隔着一个郎道。

在同时代的日本，事情也在悄悄起变化。

1937年6月，近卫文任日本首相，日本街上的妇女们都觉得生活有了希望。近卫文采取了关东军参谋长、侵华狂热分子，日本男人的骄傲东条英机的主张：鉴于中国在西安事变后，中国抗日民族统一战线日渐形成，应迅速扩大侵华战争，瓦解抵抗士气。于是，1937年8月13日，在中国上海，淞沪会战爆发，花花大世界面临挑战。而一旦上海失守，那么首都南京将洞门大开。

覆巢之下，安有完卵。1937年年底，南京失守的标志是南京大屠杀，奸淫掳掠让中国一寸山河一寸血。时为中学生的杨振宁全家辗转从合肥逃到昆明，一路上阅尽犬奔豸突的景象.....杨振宁来到昆明以后，终于安顿下来，他考上了西南联合大学化学系，后来又转到物理系。他自然没有想到，若干年以后，自己将要奔赴美国，去寻找费米。无论怎么样，当时的西南联合大学物理系并无大楼，却有一堆大师，比如周培源教授，就在莱比锡大学和海森堡打过乒乓球。这些大师也培养了一堆人类精英，杨振宁就是其中之一。这是后话。

西南联合大学的环境非常艰苦，但有一群很杰出的人物在那里，华罗庚已经从剑桥回来，他是研究素数的高手，总的来说，他喜欢把一个整数拆成几个整数的和，或者拆成几个整数的平方和，或者是几个整数的三次方的和.....这个情况，有时候是非常难做到的。数学家把这个整数分拆的问题看成无比优美的事情，其中最重要的成果是剑桥的哈代和

拉玛努杨做出的。

一个正整数拆成正整数的和有多少种方法，比如，

$$4=1+1+1+1=1+1+2=1+3=2+2。$$

哈代和拉玛努杨对整数 n 分拆的分拆种类数 $p(n)$ 写出了一个渐远表达式。——其实这个结果也可以用在群论的杨图中，是很深邃的。这个结果可以从量子统计的高温极限中推出来。

同年，费米来到了美国。这一年，当德国的哈恩和斯特拉斯曼等人发现，92号元素的原子核在中子打击下不是变成93号元素——而是像西瓜摔在地上一样分裂成大小差不多的两块……

敏感的费米意识到，类似于整数可以被分拆，大原子量的这个原子核也会分裂（可能有多块碎片），会释放出两个或者两个以上的中子，而放出的中子显然会越来越多，这样又能继续打击其他92号元素的原子核……这个过程就好像多米诺骨牌效应，意味着源源不尽的核能量会释放出来。

费米对劳拉说：“老婆，我发现了一个重要的事情。”

劳拉说：“啥事情？”

费米说：“我发现了一种中子的繁殖技术方法，这个技术方法可以制造一种异常可怕的炸弹。”

劳拉说：“我老公真牛。你想炸谁？中子也能繁殖吗？像虫子一样繁殖？还是像细菌一样繁殖？”

费米说：“繁殖系数我还没有算出来，不过，我相信在一定的条件下，中子的释放速率会高于被吸收的速率，然后这个炸弹就能实现……”

劳拉说：“我不信……”

德国的海森堡等人也意识到，92号元素能够用来制造一种威力巨大的原子弹。不过希特勒把研究导弹放在第一位，原子弹的计划，则放在第二位。海森堡相信一个伟大帝国将在自己的协助下建立起来，这个帝

国将是真正的日不落帝国，这个帝国的土地囊括了整个地球。

在这几年里，欧洲已经变了。希特勒像一个乌贼一样浮出海面。

因为欧洲局势在希特勒主张的日耳曼民族统一全地球和屠杀犹太人的思潮影响下，欧洲科学中心德国已经不适合具有自由精神的人类居住，于是，物理学的中心自觉地朝美国翕动，开始完成东学西渐的过程。

几乎所有的人都在逃命……

玻恩跑到了英国。

薛定谔跑到了爱尔兰。

爱因斯坦跑到了美国。

玻尔暂时留在中立国丹麦。

意大利作为德国的同盟，情况也是类似的，费米虽然是墨索里尼册封的院士，但他不喜欢墨索里尼。加上费米的老婆劳拉是犹太人，费米也很苦恼……1938年，诺贝尔物理学奖颁给了费米，他带着妻子和家人，先去了斯德哥尔摩……然后直接登上了去纽约的游轮。

“哦，纽约，自由女神，我费米也来了，这次我决定留下不再走……”

海森堡独自留在德国，他不知道未来怎么样，但当年研究湍流的经历告诉他，个人命运是卷在历史的湍流里的。不能活就只好死，他要好好地活下来，现在作为恺撒·威廉物理研究院的院长，终于可以走一条别人不能走的路了——他隐约觉得，藏在女人堆里的薛定谔是渺小的，我海森堡才是真的藏在花丛中的大炮。

1941年，海森堡作为一个特殊的使者访问了所有被德国占领的国家。满目疮痍令他心生快意——人生一世，草木一春，该死的赶紧去死。

第二次世界大战，打得生灵涂炭。称心快意，几家能够？整个欧洲战场，包括远东的战场，好像一块火红的电烙铁在炙烤着上面蝇营狗苟

的人群。

希特勒是一个狂热的人，他喜欢毁灭别人得到自己的快感。海森堡是为他工作的，海森堡工作的中心任务之一，就是制造原子弹。

罗马不是一天建成的，原子弹也不是一天造成的。历史的演进蜿蜒曲折。在原子弹这个事情上，首先登台的是法国人。1896年法国物理学家A.H.贝可勒尔发现铀的放射性，这是一个伟大的发现。因为伦琴发现的X射线是人工产生的，而贝可勒尔发现的放射线是天然的。

换句话说，1. 原子核是不稳定的。2. 放射性衰变是自发的。

如果原子核发生衰变，它有很多种不同的方式，在高中物理里，这一些由希腊字母表示的放射线，但通过本书的阅读，读者们已经明白，这背后的物理是量子力学里不同的衰变概率——物理学家称为“衰变道”——在这个意义上，也许这个不同的概率是可以计算的（但这个世界多数情况下是模糊的，不可计算的）。

铀及其化合物不断地放出射线，向外辐射能量。这使居里夫人发生了极大的兴趣。这些能量来自什么地方？非常幸运的是，爱因斯坦的狭义相对论马上给出了答案， $E=Mc^2$ 。

狭义相对论在理论上解释了这些能量的来源，但没有解释一件事情，那就是原子核为什么要衰变？而量子力学出现以后，人们发现，衰变是有一定的概率的，对于一些特定的原子来说，这个概率蛮大的——比如说U，就很容易衰变。

原子核会衰变，则说明原子核并不稳定。（因为丘成桐等人证明的正质量定理，以及克里斯多杜隆关于闵氏时空的非线性稳定性的证明，在经典意义上，我们的时空是比较稳定的。而前面说过，狄拉克猜想，真空在量子意义上是稳定的。所以我们暂时不用担心时空的衰变问题。）

当年在意大利科莫湖边的两个青年——海森堡和费米，今天已经渐次走到了兄弟侧目的两岸，各为其主，疲于奔命。

早在1932年，当海森堡面对原子核的时候。海森堡有一个金光闪闪的思想，换句通俗的话说：世界由男人和女人组成，男人和女人是平等的。

他想把这个思想推广到原子核里面，于是，得到了如下模型：

1. 原子核由质子和中子组成；
2. 质子和中子是平等的，它们之间通过SU（2）群在一个抽象空间中的转动联系。

他这时候的理论水平，也许在费米之上.....

（2）

1941年的某一个黄昏，一列短程火车从日本京都缓缓开出，车厢里一位沉默、戴眼镜的中年男子闭目养神，过了一会儿便摊开一本厚厚的书专心读起来。

“.....经过中川近旁，便看见一座小小的邸宅，庭中树木颇有雅趣。但闻里面传出音色美好的箏与琴的合奏声，弹得幽艳动人，源氏公子听赏了一会儿。车子离门甚近，他便从车中探出头来，向门内张望。庭中高大的桂花树顺风飘过香气来，令人联想贺茂祭时节。看到四周一带的风物，他便忆起这是以前曾经欢度一宵的人家，不禁心动.....”

车上这个正在看黄色小说《源氏物语》的中年男子叫汤川秀树，是京都大学的物理学教授，他正在下班回家的路途上。他的计算表明，原子核里面有一种巨大的力量把质子和中子束缚在一起，这个强大的信使粒子在质子和中子之间就好像一个介绍人，他称为介子，介子的存在使得原子核基本保持稳定，也使得原子核这个小宇宙内的温度为介子的质量（温度与能量通过玻耳兹曼常数相互联系起来），大约200个电子质量（在物理学中，质量、能量和温度，都可以认为是同一个意思）。可惜他的计算没有表明这个巨大力量将进入日本的未来光锥：他的同胞将

在广岛接受死亡之吻。

1941年12月7日上午，京都的天空有点阴郁，战争的阴霾弥漫在岛国的上空。

物理系教授汤川秀树放下手中的书，仰望灰蒙蒙的天空，眼神有些倦怠，想起了1926年自己刚考上大学时候的情境：那时候的他，年轻英俊，朝永振一郎（Sin-itiro Tomonaga）是自己的同班同学，他是20岁，自己19岁，大家一起看川端康成写的《伊豆的舞女》，薰子经常和自己一起手拉着手躺在19岁的床上.....

今夕何夕？

汤川觉得有点sentimental，他也很想知道，有着丰厚漆黑的秀发和像鲜花娇美的面孔，眼角处涂抹着古色胭脂红的薰子，现在何处？19岁的邂逅和告别，告别也就是永别，再来也许要在天上团聚。

汤川正在遐想之中，外面进来一个人告诉了他一件可怕的事情：今天清晨，大日本帝国海军的航空母舰舰载飞机和微型潜艇突然袭击了美国海军太平洋舰队在夏威夷基地珍珠港以及美国陆军和海军在欧胡岛上的飞机场.....

汤川秀树是一位没有到过欧美留学，而是在日本本土生土长起来的理论物理学家。他有很严重的自卑情绪。他害怕地说：“这下完蛋了，我们日本招惹了一个巨人。”

日本在正式宣战之前，就偷袭了珍珠港，引起了美国全社会的同仇敌忾，狮子终于要发怒了。过了4年，专门为日本量体裁衣定做的原子弹就造了出来。

当德国和美国同步进行原子弹的研发，理论物理学家海森堡在技术上的缺点暴露无遗（比如他不认为石墨是有效的减速剂，只能用重水，但重水很难找到），而且令人意外的是，他还算错了原子弹的临界质量，以为大约需要几吨高纯度的浓缩的 ^{235}U ，实际上只需要18到40千克左右。而费米等人，则开始了技术上对炉火纯青的摸索。1939年的时

候，当玻尔坐船从丹麦赶到纽约，把原子核在中子撞击下裂开的消息传递到美国以后，几个在美国的匈牙利人坐不住了，他们是维格纳、西拉德等人，他们渴望联络爱因斯坦，给罗斯福总统建议，抢在德国之前制造原子弹。在这之前美国的惠勒和玻尔用量子理论计算了一下，他们想判断一下 ^{235}U 和 ^{238}U 吸收中子以后哪一个更加容易发生分裂，他们的计算结果是找到了一个关键性指标：吸收中子后的原子核的电荷的平方与质量之比率，比率大容易裂变。 ^{235}U 的该指标比 ^{238}U 的大，所以更容易裂变。第二个因素是原子核中如果有偶数个中子，就会变得更不稳定，所以 ^{235}U 吸收一个慢中子以后变成 ^{236}U ，原子核内有偶数个中子，就会变得不稳定。结合这两个因素， ^{235}U 适合做原子弹。但 ^{235}U 在天然U矿中的比例大概是0.6%，99%都是 ^{238}U ，因此，需要提炼。

在技术层面上，需要有3个主要的考虑：

1. 把 ^{235}U 从 ^{238}U 堆里分离出来——这类似于把铁粉从铝粉中找出来。
2. 实现中子的繁殖——首先中子必须减速，因为速度大的中子，波动性小，撞击面就小。而中子速度减小以后，波动性变大，就好像一个乒乓球放大成了一个篮球那么大，容易撞上原子核。其次是中子不能被环境强烈吸收，中子数目就好像混沌动力学中的虫口模型，对环境非常敏感，对于虫子来说，模型表明，环境的微小改变将引起虫子数目的确定性混沌——当然，这又被称为逻辑斯蒂模型，仅仅是模型而已。对于中子也是差不多的，很难计算真实情况。
3. 制造设计核反应的炉子——如何安放 ^{235}U 块，计算出临界的质量——放置镉棒可以吸收中子，使得核反应可以控制，至少别炸死实验人员。

费米从纽约哥伦比亚大学被集中到芝加哥大学以后，他领导的小组马上实现了受控的核反应。

接着，美国动用全国的人类精英，开始在洛斯阿拉莫斯的寒冷地带

制造实验原子弹。

“二战”打得斗转星移。

物理学的发展也停滞了，一切都要为战争服务。

1945年，苏联红军朝柏林推进，海森堡在研究所里透过玻璃看到惨淡无光的太阳。看到外面是荷枪实弹的警卫，他们奉命可以枪杀任何一个擅自离开岗位的研究人员。

可是，海森堡已经等不及了，因为，如果他再不逃走，也许会死在苏联红军的手里。他不想死在斯大林的手上，于是，决定赌一把。

1. 不走，可能死在斯大林手上；

2. 走，可能死在希特勒手上。

他选择了走。

他下楼，朝自己的自行车走去……

抬起屁股，他上了车，脚蹬子显得很沉重……

他想骑车赶到慕尼黑的家里去，前面是一个岗哨……

缓慢地移动，离铁丝网越来越近……

“站住！”警卫在背后用枪顶住了自己的腰……，“干什么去？！”

海森堡连忙下车，从上衣口袋里掏出了一根香烟说：“哥们儿，抽烟……”

士兵把枪从腰上移开，接过了烟，打量了一下这个教授说：“老师，您要过去呀……”

海森堡用有点颤抖的手打着了火，给士兵点着了烟，说：“是啊，我过去有点事情……马上回来……”

士兵吸了一口烟，用慈悲的眼光看了他一眼，然后说：“神爱世人……你去吧。”

海森堡慌忙上了自行车，连谢谢也忘了说，飞也似的逃走了。不久，他被从诺曼底登陆的那帮盟军俘虏了。

海森堡傻了，心想，妈的，刚出虎穴又进狼窝。上帝啊，你消遣我

吧？

海森堡被带到了英国的监狱里，他这才意识到，自己在计算的时候漏掉了另外一个可能性。

3. 走，死在丘吉尔手上。

(3)

1945年8月6日8时15分，美军一架B-29轰炸机飞临日本广岛市区上空，投下一颗代号为“小男孩”的原子弹。“小男孩”是一颗铀弹，长3米，……

“小男孩”代表上帝和珍珠港死难者亲吻了广岛。

你出现，像一盏灯。

燃烧了，我的瞳孔。

19 原子弹研发与物理学熊市

(1)

1939年1月16日，星期一，纽约的哈德逊河码头繁忙依旧。一艘名叫MS Drottningholm的轮船正在缓缓停靠向码头，岸边已经站着一个前额高高的意大利人，正在焦急地等待，他的身边站着一个美貌的女子，皮肤黝黑，眼神顾盼神飞，这个女人的身边站着两个孩子。他们正是费米一家，正在这里等待玻尔。

在这一家人的不远处，还站着一个名叫惠勒的美国青年，他也在这里等待量子力学教皇玻尔访问美国。玻尔这次到美国的主要目的是与普林斯顿高级研究所的爱因斯坦探讨铀裂变问题。玻尔也是在1月7日才刚刚听说原子核分裂实验这个事情，当时他与他的儿子爱瑞克正在哥本哈根准备要坐火车去港口城镇古腾堡搭乘MS Drottningholm号轮船。当时有一个从德国移民到丹麦的年轻物理学家，此人的姨妈物理学家迈特奈（Lise Meitner）告诉他一个消息：德国的化学家哈恩与斯特拉斯曼在柏林的实验室里用中子去轰击铀，却产生了钡元素。哈恩写信给迈特奈描述了这个实验现象，迈特奈于是把这个消息告诉了她的外甥，并且认定这个现象的发生是因为铀原子核发生了分裂。而这个外甥把消息透露给了玻尔。

于是，玻尔带着这个秘密来到了美国，他同时把这个事情写了一篇物理论文投稿给了《自然》杂志，后者在2月就刊登了这个重要的文章。同时他和费米在华盛顿大学举行的一次理论物理学会议上交换了各自的研究心得。在这次交谈中，关于链式反应的概念开始成型。所谓链式反应，就好像点火柴一样，以前面的燃烧产生的热量来启动后面的燃烧过程，类似于一场雪崩。是年3月费米等人进行实验以确定铀核裂变

释放出的中子数目到底是几个：实验结果表明，铀核在裂变时能够释放多于两个的中子，因而铀原子核一个接一个分裂的链式反应应该是可以实现的。至此，在理论上能否实现核分裂链式反应的问题已经得到基本解决。由于纳粹德国也在沿着这一方向进行研究，聚集在美国的各国著名科学家们强烈地预感到，美国政府应该利用这一最新科研成果，开始研制一种威力强大的原子武器，而且必须赶在德国人前面。1941年7月，费米等人在哥伦比亚大学，开始着手进行石墨-铀点阵反应堆的研究，确定实际可以实现的设计方案。

(2)

1941年9月，经过一层又一层的审核，纳粹终于同意海森堡去哥本哈根讲学一周。当然海森堡有一个更重要的任务在身——就是去拉拢刚从美国回丹麦不久的玻尔。这一次会面的真实场景，如话剧《哥本哈根》所讲述，没有人知道了。

玻尔老师和海森堡曾经情同父子，在量子力学创立的过程中两人都作出了杰出的贡献。因此当海森堡在四十岁的不惑之年，再次来到了哥本哈根大学，穿越那些熟悉的研究楼，走过熟悉的饭厅。他的情绪也是非常复杂——因为玻尔的祖国丹麦正被海森堡祖国的军队占领着。

那天晚上，海森堡又来到了玻尔家中，与他长谈。海森堡期望像当年讨论量子力学那样彼此没有什么隔阂，但是时间再也回不去了。玻尔的表情凝重，因为德国人侵略了他的祖国，在全欧洲大肆地迫害着犹太人。

海森堡说：“老师，我们的坦克已经兵临莫斯科城下，英国也快被我们的飞机炸平了。您应该加入我们，研制出核武器，这样战争就会很快结束了。当年玻恩教授跑去英国的时候我也被牵连，那时候我没有能力保护他，但是我现在身居要职了，我会用我的身份保护您以及其他的

犹太物理学家。尽管爱因斯坦和薛定谔都走了，柏林还有普朗克先生和劳厄（普朗克的学生，因发明X射线衍射技术获得1914年诺贝尔物理学奖）这样的您的故交，所以请您考虑加入到我们这个阵营吧。”

玻尔沉默了片刻说：“我们只谈物理，不要谈这该死的战争，但既然你说了，我就告诉你我的真实想法。在我看来，这场战争对于你们德国人是复仇般的宣泄，对我们丹麦是彻头彻尾的灾难，我不会选择与侵略者合作。”

海森堡说：“我和我的妻子伊丽莎白不敢想象我们的孩子像我小时候一样饱受战争失败的折磨，我希望德国获胜，越快越好，因此我会全力以赴地为国效力，发展核武器。”

玻尔严厉地说：“好吧，那我们已经无话可说！”

以上是他们会面的一种猜测。

1941年的12月6日，也就是日本偷袭珍珠港的前一天，罗斯福总统下令设置专门机构，以加强原子能的研究。此时，康普顿被授权全面领导这项工作，并决定把链式反应堆的研究集中到芝加哥大学进行。1942年年初，哥伦比亚小组和普林斯顿小组都转移到芝加哥大学，挂上“冶金实验室”的招牌。这就是后来著名的国立阿贡实验室的前身。在芝加哥大学的这个“冶金实验室”里，费米所领导的小组主要是设计建造反应堆。在建造并试验了30个亚临界反应堆实验装置的基础上，最后才制订出建造真正反应堆装置的计划。1942年11月，这个反应堆主体工程正式开工。由于机制石墨砖块、冲压氧化铀元件以及对仪器设备的制造很顺利，工程进展很快。费米的两个“修建队”，一个由津恩领导；另一个由安德森领导，几乎是昼夜不停地工作着。而由威尔森所领导的仪器设备组，也是日夜加班，紧密配合。反应堆一天天朝着它的最终形象增长。为它工作的人们，神经紧张的程度也在增加。虽然从理论上说，他们明白：在这反应堆里，链式反应是可以控制的。但毕竟是第一次，是不是可控还需要用实践来证明。

(3)

也就在这研发原子弹的过程中，物理学也走进了熊市。

1927年物理学的高潮过去以后，到第二次世界大战开始，这中间有10年的黄金时间，到了“二战”时期，物理学也转为战争服务，基本没有大的进展。爱因斯坦和诗人薛定谔依然不相信哥本哈根学派的解释是完备的，在他们看来，量子力学不是一个原理性的理论，而是一种以人为中心的类似于托勒密地心说的理论。可是，地球并不是星星们的中心啊……虽然爱因斯坦一直不相信量子力学是对的，也提出了光子箱来轰击不确定性原理，但这个思想实验又被玻尔用广义相对论化解。在这中间的黄金10年中，唯一值得说的进展，则是狄拉克对磁单极子的研究。

1930年7月31日，一个19岁的印度人，踏上去英国的航程。在那10多天的海上漂流中，钱德拉塞卡把自己藏在下等舱里，躺在床上撅着屁股演算一个关于引力坍塌的问题。

同船的其他人沉浸在美酒、音乐和舞蹈的狂欢中，白矮星的质量上限却被这个少年人的计算最后确定为1.4个太阳质量。继16年前的拉玛努扬之后，印度将再次有人要在剑桥独自舞蹈。钱德拉塞卡到了剑桥大学以后，师从富勒读博士学位——他成了狄拉克的师弟，也算是张宗燧王竹溪的师兄了。

钱德拉塞卡的研究结果就像一本《印度爱经》，西方人看了以后是惊呼怪异。著名的天文学家爱丁顿教授认为，钱德拉塞卡关于恒星引力坍塌的结果，肯定是错误的，因为如果自然界不存在其他力量抗拒恒星的引力坍塌，那么，很明显，巨大质量的恒星会在空间引起一个奇怪的黑洞……

1931年的一天凌晨，狄拉克洗刷完毕后，虽然天才蒙蒙亮，但他决定去大学办公室开始工作了……到了大学，在楼道里刚要开门进办公

室，他看见身边有一个鬼鬼祟祟的印度人，腼腆得很，这个人手里拿着一本书，朝自己诡异地笑了笑，他笑起来有点像释迦牟尼，这让狄拉克感觉很害羞，慌忙也笑了笑，准备掏钥匙……

这个印度人却走了过来……

“早上好，狄拉克博士。”这个印度人过来搭讪了。

“……”狄拉克感觉自己的嘴巴动了一下，却没有声音。

“狄拉克博士，你的讲义《量子力学原理》和牛顿的《原理》一样，都是经典。”这个印度人接着说。

“……”狄拉克抓起了钥匙把手从裤兜里拿出来。

“狄拉克博士，我是从印度来的留学生，我叫钱德拉塞卡……”印度人还是在说话。

“有事？”狄拉克把钥匙插了锁孔，终于微弱地说了一句。

“没有事情。打扰了——您觉得根据广义相对论的引力坍塌以后，会不会引起空间上的一个奇点？一个黑色的洞？”印度人赶紧抛出了一个学术问题，希望打动狄拉克。

“引力坍塌？球对称的吗？不过你好像问错人了，我不是干这个的。”狄拉克这下果然被激活了，多说了三两句，“你的问题很好。”说完，他扭转了一下钥匙，门开了，他进去以后，转身就把门关上了。钱德拉塞卡愣在门口，半天才悻悻地离开，离开的时候他的脚步软绵绵的那么虚弱，好像是空气在推着自己前进。

狄拉克脱掉外套，坐在办公室的椅子上，刚才那个黝黑的印度阿三问的问题却浮上心头。如果大质量引力坍塌真的会形成一个黑洞，那么，人们该如何探测这个黑色的洞？

很显然，这个黑洞会产生引力场，那么，应该可以从引力场的数据中读出引力源（黑洞）的性质。可是，怎么读出来呢？——这问题依然类似于听见鼓的声音来分辨鼓的形状。这个反问题可不简单（黑洞作为一个引力源，可以通过引力场的曲率张量构造一些所谓纽曼-彭罗斯常

数来标记引力源的性质），狄拉克也不能超越时代，他想了半天，觉得自己不懂引力，算了，还是想点量子力学的问题。

好，如果用磁场来代替刚才那个引力场，情况会怎么样？

狄拉克做物理向来是风烟俱净清爽透明的。他的笔在草稿之上从流飘荡，任意东西。磁场如果从一黑洞（磁单极子）发出来，标记为**B**。磁场**B**是一个矢量场。

狄拉克马上下意识地吧磁场**B**写成了另外一个矢量场**A**的旋度——虽然他不清楚这样搞是不是真的对。这样，他就有了一个磁单极场，这个磁场**B**跟点电荷的库仑场**E**很类似，它们的散度都正比与原点的狄拉克 δ 函数。但区别在于，在磁单极场中，**B**是另外一个矢量场**A**的旋度，而是库仑电场中，电场**E**是另外一个标量场的梯度。



狄拉克的墓碑

问题马上就很明显，在磁单极场中，矢量场**A**在球坐标系里怎么也写不完全，在球面上至少有一个点是没有定义的。在整个三维空间中看到，这些矢量场**A**没有定义的点组成了一个从磁单极出发延伸到天空之外的一条弦。（这是数学上没有定义的一条弦，称为狄拉克弦，在物理

上可以通过引入别的坐标系取消，这不是物理奇性，只是坐标奇性。)

狄拉克马上就把量子力学也放了进来，考虑一个电子绕着磁单极子走了一圈。因为电子是不应该撞上狄拉克弦的，但狄拉克弦在空间中如果真的存在，将破坏空间的各向同性，所以，狄拉克弦不应该真的存在。狄拉克引进了另外一个坐标系，把狄拉克弦取消了。

那么，在两个坐标系里，矢量场 \mathbf{A} 可以分别定义。

在两个坐标系的重叠区域，任意一个封闭曲线上运动的电子，它的波函数的相位变化应该是不依赖于坐标系选择的。这样很容易就得到：

$$eg = n$$

也就是说，电子电量和磁单极子磁场量子的乘积正比于所有整数。

换句话说，假如

1. 存在磁单极子；
2. 磁单极子磁场可以写成一个矢量的旋度；
3. 空间是三维的。

那么，电荷必须是量子化的。

狄拉克做完这些以后，站到窗口，他看见太阳还在半空中，这是早晨八九点钟的太阳。晨曦之下，刚才那个印度人正在树下拿着一本书，用短小的铅笔头在上面做什么计算.....难道那人是在计算引力场单极子吗？

狄拉克陷入了遐思。他感觉理论物理学的很多问题已经被解决了，自己也正在变老。也许，物理学的熊市要来了！

20 枪手，猫论

(1)

1933年的一个夜晚，荷兰。

性都阿姆斯特丹的凄迷灯光照亮了街道，大雨滂沱已哭成一个泪人，蓝色的油纸伞下，一个50多岁的小矮个子步履有些蹒跚，他把伞压得很低，路人无法看到他的那双带着雨水的迷离眼神.....他好像踟躇不前，似乎有什么心事，只有他自己知道，他的裤兜里有一把左轮手枪。

“喂，进来！”路边妓院里的小姐冷冷地打招呼。

埃伦费斯特教授偷偷地看了一下妓院门口的花灯，艳丽得像在流血。她在朝他勾手。埃伦费斯特把脸转向地面，看见地上雨水冒泡，好像是一个饥渴的路人在喝水一样.....埃伦费斯特咽下一口唾沫，他下意识地摸了一下裤兜里的左轮手枪，步履匆匆朝下榻的旅馆走去.....他的小儿子正在旅馆里。

埃伦费斯特悄悄地打开了旅馆的房门，看见房间里凌乱不堪，自己的小儿子正一丝不挂地站在凳子上仰望着天花板。小儿子低头用发散的瞳仁看了埃伦费斯特背后的墙壁一眼，骂道：“畜生！”

埃伦费斯特没有应声，因为他的这个儿子有精神分裂症，脑电波是一根比较平坦的直线，不像正常人是一根混沌起伏的曲线，因此打人骂人是经常的事情。

小儿子从凳子上跳下来，踹了埃伦费斯特一个屁墩，一下就把埃伦费斯特踢倒在地上，他高声嚷道：“call me god！”

埃伦费斯特躺在地上，裤兜里的枪也掉了出来，他拿起枪。

“上帝啊！主啊！我埃伦费斯特到底做了什么孽！”埃伦费斯特把枪对准了儿子的肚子，眼泪和脸上还没有干的雨水似乎夹裹了整个天地。

埃伦费斯特从地上爬了起来。

突然，他转身把床上的棉被拿起来，猛地裹住了儿子的脑袋。

枪声响了。



埃伦费斯特所拍摄的照片：玻尔和爱因斯坦在埃伦费斯特家里，左为玻尔

他的儿子头颅中弹，躺在了血泊之中.....

埃伦费斯特跪倒在地上，看见儿子已经死了，抹了一把眼泪，用颤抖的手把枪口对准了自己的太阳穴。

在临死之前，他多么想再看一下这个花花世界，从维也纳走到阿姆斯特丹，自己花了一辈子的时间，命运起伏跌宕，多么不易。

但现在时间已经不多，他要解脱了。他想起的第一个人，是自己年轻时代的老师玻耳兹曼，那个自杀的统计物理学家——当时他在维也纳大学上课的情景历历在目，他跟随玻耳兹曼学习热力学中的分子运动论。

外面漂泊的大雨正在打击窗户，上帝似乎正在窗外窥探，埃伦费斯特朝窗口打了一枪。

玻璃落了下来。

埃伦费斯特朝自己的太阳穴开了一枪……世界安静了。

埃伦费斯特的死在物理学界一石惊起千层浪，连爱因斯坦也暗地里淌泪。爱因斯坦还能记得自己当初在德西特教授的家里和埃伦费斯特开怀畅饮的情境。德西特解出了爱因斯坦引力方程的一个解，这个解能够描述的宇宙的尺寸随着时间指数膨胀——这个宇宙空间是运动的，而且是真空的，遥远的物体所发出的光的波长会被拉长，颜色会变红——后来这个宇宙模型被埃伦费斯特的一个叫弗里德曼的俄国学生推翻，后者也得到了一个膨胀的宇宙解，这是后话。埃伦费斯特带的两个学生发表了电子自旋的文章。虽然爱因斯坦也被这些事情搞得稀里糊涂的，刚体才会旋转（刚体自转的运动一般情况下运动方程是解不出来的，被称为不可积的），而电子好像是一个没有大小的质点，怎么会自旋呢？但也许这就是荷兰的物理学家最杰出的工作了。

（2）

早在1913年的秋天，斯塔克（Stark）的发现在物理学家中间引起了另一次轰动：他发现了电场对氢谱线结构有影响。卢瑟福从普鲁士科学院收到斯塔克的论文后，立刻就给玻尔写了信：“我认为时机已到，您现在应该针对塞曼效应和电效应写一些东西，如果有可能将这些效应和您的理论调和一致的话。”

玻尔响应了卢瑟福的挑战，尝试着仔细考察了这个问题，而且很快就弄清楚，在电场的效应和磁场的效应中，必须处理两种很不相同的问题。

关于塞曼（Zeeman）效应在1896年就有洛伦兹和拉摩尔的解释。而所谓反常塞曼效应的出现却带来一些新的疑难；只有在10多年以后，当线系谱中各谱线的复杂结构被迫根到一种内禀性的电子自旋时，这些疑难才算得到了解决——这里面也有埃伦费斯特的功劳，没有他的支

持，他的学生是不敢发表这个在当时看起来离经叛道的假设的。然而，在电场的情况下，不能预期谐振子发射的辐射正比于电场强度的效应，从而斯塔克的发现就肯定地排除了将电子的弹性振动看成线光谱的起源那种通常的见解。但是，对于电子绕核的开普勒运动来说，即使是比较弱的外电场也会通过久期微扰而使轨道形状及轨道取向发生相当大的变化。通过研究轨道在外场中仍为纯周期性的那种特例，有可能应用和适用于未受扰氢原子之定态的那种论证类型相同的论证，来推求斯塔克效应的数量级，特别是解释该效应在氢光谱线系中从一条谱线到下一条谱线的迅速增大。但是，这些考虑很显然地表明，为了解释现象的更精致的细节，原子体系的定态分类方法还是发展得不够的。

如1916年索末菲（Sommerfeld）在他的著名论文中所证明的，角动量和径向运动作用量的分别量子化，可以详细地解释所观察到的氢原子光谱和氦离子光谱中各谱线的精细结构了。索末菲通过熟练地应用多周期体系的量子化方法，他们已经能够完全符合于观察结果地通过并合而得出氢谱线的分解的谱项。这种方法和埃伦费斯特为了适应热力学的要求而在1914年表述的定态浸渐不变性原理是相容的：各个量子数按照古典力学的作用量积分，并不会因外场的改变而改变，如果这种外场的改变比体系的特征周期变化很慢的话。

虽然埃伦费斯特有不少在物理学上的贡献，但总的来说，埃伦费斯特似乎依然是一个不被世人理解的人。

泡利与埃伦费斯特是朋友。在他们第一次见面时，埃伦费斯特说：“我喜欢你的物理胜过喜欢你本人。”

泡利说：“我的感觉恰好相反。”

其实，埃伦费斯特是一个真正的传奇。埃伦费斯特发现了一件很重要的事情：如果量子力学真的是万能的，那么，地球是围绕太阳公转的一个在椭圆上运动的波函数。但埃伦费斯特发现，这个波函数必然会扩散到全空间——也就是说，量子力学描述中，地球作为一个波函数不可

能不爆炸开来。这就是著名的埃伦费斯特扩散。

(3)

红尘任他凄凉。

埃伦费斯特永远离开了。但埃伦费斯特的问题留了下来：如果量子力学是万能的，那么地球是一个波函数，根据埃伦费斯特扩散，随着时间演化，地球要在全空间弥散开来——因为地球处于太阳的万有引力场中，而这个引力场是平方反比的牛顿引力，因此，地球的哈密顿算子下波包的演化必然要扩散。

但地球现在一直好好的，并没有扩散到全宇宙空间，因此，量子力学不可能应用到宏观的天体力学问题上，这简直已经是显然的。量子力学在微观世界是一个很好的描述，为什么在宏观世界就不行了呢？到底哪里出了问题？埃伦费斯特走了以后，没有人再思考这个问题。因为冯·诺伊曼写了一本量子力学的书，论证说量子力学在数学上是完备的理论。（后来才被发现，冯·诺伊曼犯了一个很低级的数学错误，那就是他居然认为两个算子和的期待值等于期待值的和，其实 $\langle A+B \rangle = \langle A \rangle + \langle B \rangle$ 对于A，B不对易的时候并不成立。）

薛定谔深刻地感受到了其中的悲哀。他的脑子也有点乱掉了，他无法看到那些乌云之上的阳光——作为一个诗人，他能看到的量子力学是一片无边的荒漠里的一个孤坟，早已经死亡，独留青冢向黄昏。

只不过他也并不确定量子力学能不能应用到宏观物体，于是，脑子很乱的他一心只想着把事情搞砸，让玻尔他们下不了台阶。于是，在1935年，也就是埃伦费斯特尸骨未寒的两年里，他得到了另外一个版本的故事——薛定谔的猫论。薛定谔的猫基本阐述是这样的：“设想有一个箱子，里面有一只活猫。一个装有镭的容器及一个装有氰化物的小瓶也放在箱子之中。镭原子会发生衰变。在这个装有活猫的密闭的箱子

里，如果镭发生衰变，会打碎瓶子，使氰化物从小瓶之中释放出来，从而杀死猫；如果镭不发生衰变，小瓶也不会破碎，猫会活下去。按照哥本哈根解释，在打开箱子看猫的死活之前，猫既是死的，也是活的，因为两种可能性都存在。而且，箱子中的猫会保持这种既死又活的状态，直到有人打开箱子，发现猫要么是死的，要么是活的为止。”

其实，猫这样的宏观物体根本就不可以简化为一个波函数来描述，在量子世界里，猫与地球一样，都是很大很大的宏观对象。

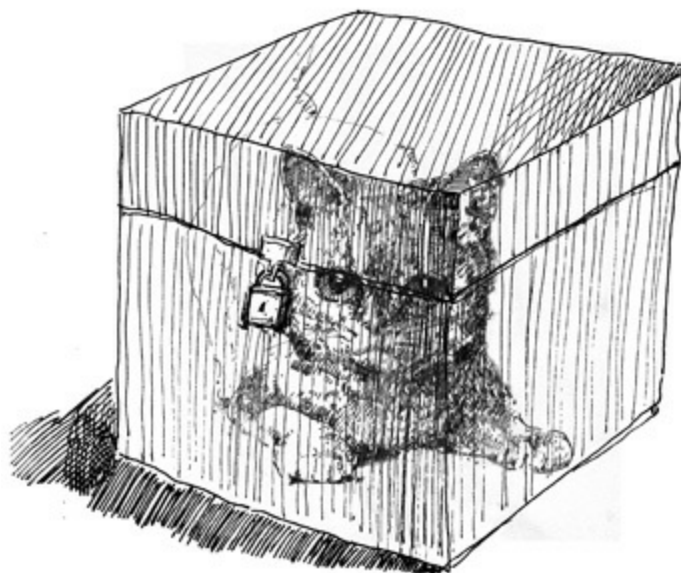
不过，薛定谔也不是傻子，他这样搞只不过是吓唬吓唬玻尔、海森堡那些他心目中的庸俗量子物理学家，因为后者声称量子力学太完美了。薛定谔是一个诗人，审美与别人迥异，别人都说一个东西好的时候，他一定会觉得这个东西是庸俗不堪的。

薛定谔把猫从他的笼子里放出来，果然很有效果，这个猫像疯狗一样咬人，连海森堡也不知道怎么用哥本哈根解释来描述这个疯猫了。

薛定谔太高兴了，因为他知道海森堡他们已经在量子力学道路上走火入魔，根本不可能意识到，猫这样的宏观物体，也许不可以用波函数来描述的。海森堡之流，就好像是《天龙八部》里的纠摩智，为了练六脉神剑，已经得了抑郁症。真是可笑啊。

薛定谔早已经构造了一个不扩散的波包，不过不是在平方反比引力场中，而是对于弹簧振动，他可以构造出一个量子力学波函数（称为相干态，在空间上具有高斯分布的形态，在经典物理中，孤立子就是不扩散的波包，这其实是一种扩散和凝结相互平衡的能量状态），这个波函数描述的弹簧振动确实有一个不扩散的波包。

薛定谔做完这些，感觉无人喝彩，空虚彻入骨底。他也无所谓，转身去找他的情人了。



薛定谔的猫 绘画：张京

21 在战后的废墟上

(1)

到了1942年，薛定谔已经不仅仅沉迷于乳房研究，他已经老了。对自己的一辈子有了很好的评价：

重剑无锋 大巧不工 四十岁前恃之横扫天下

四十岁后 不滞于物 草木竹石 皆可为剑

作为一个55岁的人，他已经可以拿起草木竹石当武器了，他不想再使用重剑（薛定谔偏微分方程），但他发现，简谐振动的能级可以用代数解答写出。什么叫代数呢？

就是没有微积分的数学演算。比如因式分解（草木竹石）。

$H = X^2 + P^2$ 可以被因式分解为

$$H = (X + ip)(X - ip)$$

其中 $i = \sqrt{-1}$ 。他换了一个记号，记

$$A = X + ip$$

解出了能级以后，还可以检验 $A|乳房\rangle = a|乳房\rangle$ 成立。在这里的乳房，就是一个永不扩散的波包。

A被称为湮灭算子，因为这个算子对真空态的作用为0。

(2)

在“二战”之中，这一点已经快被忘却。

战争终于快结束了，希特勒也快消失了。在美国，有一个纽约客，在一个野鸡大学读书，天空整天灰蒙蒙没有未来，这个人叫施温格。他非常神奇，数学水平很高，动辄就能写出300页的数学论文，他没有得

到学士学位之前就已经写完了博士论文，没有人怀疑他比他的同学费曼要聪明一些，简单地说，施温格是量子电动力学的奠基人之一。他后来能模仿薛定谔的手法，能把泡利的自旋角动量SU(2)代数也用类似的方法实现出来。

随着战争的结束，渐渐地，物理学的熊市正在复苏，有了一些量子电动力学的萌芽。量子电动力学是战后的新一代人建立的，他们是朝永振一郎、费曼、施温格（Schwinger）、戴森、朗道。施温格和费曼一起发现了量子电动力学。已经说过，施温格十几岁时混得并不好，就在纽约一个社区大学混日子，但有一天著名物理学家拉比（I.I.Rabi）和另一位物理牛人在谈论量子电动力学的问题，施温格插话了，“这个少年尖锐的发言结束了这场争论”。拉比托关系把他招进纽约的哥伦比亚大学，从此施温格结束了他的山寨物理学生活，开始了顺风顺水的学术研究……施温格以超强的算功闻名。有段时间，施温格在奥本海默那里干活。有两个年轻的物理学家来找奥本海默请教一个问题。奥本海默就告诉他们这样这样算，回去自己算吧。当天施温格狂算了一个晚上，最后就把结果写在一破纸上，塞到一口袋里。过了五六个月，那两个人回来了，高兴地把结果拿给奥本海默看，奥本海默就对施温格说你不是早就搞定了吗，你去对一下。施温格就回去把所有的衣服翻了个遍，终于找到了那个小纸团。对了一下，回来告诉大部分是对的，只差了一个因子。施温格的工作时间和其他人是正交的。一般他下午5:30去办公室，那时候别人都回家了。所以如果有问题，总是留个条子在他的桌上。一次一个物理学家不太懂贝塞尔函数，就留个条子在他的桌上。第二天这个物理学家发现一本40页的答案。但是他发现这个结果没有物理意义。就又留条子问施温格你是不是弄错了？施温格自信地说没问题。这哥们只好自学相关的数学，好长时间自己也弄出来了。再对比了一下施温格的答案，发现施温格用错了一个公式。这事很丢面子，于是施温格恼火地说：“下次再也不到书上抄公式了，用什么公式之前一定自己先推导

一遍。”

朗道等人在莫斯科发现，对于可重整（就是可以想办法把无穷大发散的问题通过掩耳盗铃的方式忽视）的量子电动力学来说，也存在其他的困难，那就是相互作用的耦合常数随着能量的增加而变大的过程之中，在能量大了一个有限数值的时候，耦合常数就已经是无穷大了。这个时候的能量，被称为朗道极点。换句话说，在这个有限能量之上，量子电动力学就会失效。因此，反过来也可以说，朗道极点以下能量范围内的量子电动力学，是一个物理上有效的理论。耦合常数依赖于能量这个观念是重整化群的基本观念。耦合常数表示的是相互作用的强度，就好像是一男一女两个人相互交往的亲密程度的表征，这个亲密程度并不是一个常数，一开始的时候，能量比较低，男方一般会送女方玫瑰花，女方也会送男方巧克力之类，这时候的花费大约在100元这个数量级上，但随着两人亲密程度的增加，所需要的能量也在上升，最后发展到买钻戒，需要大概10000元了，而关系最亲密的时候，则需要很高的能量，那就到了谈婚论嫁的时候，这个时候所需要买的房子，大约需要1000000元了。朗道他们所发现的极点说明了一个现象：只需要有限的钱，这男女的亲密关系就已经到了无穷大的程度了。

这在物理上是有点说不过去的。

外一篇 早期量子场论简史

在1926年，在关于点粒子的波动力学刚刚建立之后，出现了O.克莱因，B.A.福克和W.高登的致力于波动方程之相对论推广的文章。克莱因利用了五维抽象空间的方案。在福克的文章中，五维空间也起了同样的作用。1927年，狄拉克发表了一篇《辐射的发射和吸收的量子理论》的著名论文，首先对原子与辐射场的相互作用进行系统而精细的研究，表明通过薛定谔方程不仅可以把量子力学用于原子，而且还可以用于辐射

场。

那么，如何把一个场量子化呢？

根据德拜在1910年提出的设想，狄拉克把自由空间中的电磁场当作一种量子化的振动过程。于是，电磁场可以按本征振动模式作傅里叶分解为无穷多个不同频率的简谐振动，每种简谐振动模式都具有一定的波矢量、频率和偏振方式。在自由空间，电磁场可以看作是无穷多个没有相互作用的谐振系统，每一个谐振子都对应于一个满足薛定谔方程的本征振动模式。

1928年，约旦和维格纳在研究费米子场的量子化时，将这种方法又应用到电子上，按上述要求提出了电子场的量子化方案。同年，狄拉克推广了约旦等人的技术，提出了一个用谐振子系统表述电磁场的量子化程序，引入了量子场论中的产生与湮灭算符，以及二次量子化方法。

随后不久，约旦指出：狄拉克的产生与湮灭算符技术不仅适合于光子一类的玻色-爱因斯坦系统，而且适用于费米-狄拉克系统。

量子场论实质上是无限维自由度系统的量子力学。它给出的物理图像是在空间充满着各种不同物质的场，它们相互渗透并相互作用着。场的激发态即为粒子的出现。不同的激发态，则表现为粒子的数目与状态的不同。场的相互作用又可以引起场激发态的改变，这就表现为粒子的各种反应过程。在考虑相互作用后，各种粒子数一般不守恒，因此量子场论能很好地描述原子中光的自发辐射与吸收，以及粒子物理学中的各种粒子产生和湮灭的过程。

现在，我们转向曾使量子电动力学遭受灭顶之灾的无限大问题。

首先美国原子弹之父奥本海默证明：用形式上协变与自洽的海森堡-泡利理论，来计算原子能级的相对论修正项时，会得到正比于

$\int_0^\infty \frac{dk}{k}$ （对数发散）的无穷大。而当计算电子在自场中的自能（至第

二级) 时, 也得到一个无限大的结果: $\int_0^\infty k dk$ (二次发散)。

对于发散问题的系统研究, 始自20世纪30年代中期。狄拉克与海森堡证明: 空穴理论中的某些无限大可以一致地加以消除, 以产生一个形式上洛伦兹不变的理论。海森堡所采取的方法就是对电子的电荷“重新定义”。

粒子自能无限大的问题在经典的洛伦兹理论中就出现了, 因此洛伦兹给电子指定了一个有限的半径。

韦斯科夫证明: 空穴理论的自能只是对数发散, 这种发散存在于精细结构常数 α ($2\pi e^2/\hbar c \approx 1/137$) 的一切级别中。这就是说, 进入更高级项不会产生新的发散。贝特在这基础上提出了质量重整化程序, 而且推广到具有相对论效应的场合中。自由电子的电磁质量, 和电子本身所引起的场的能量成正比; 这种场就是虚光子云。束缚电子的电磁质量可能是无穷大, 因为当电子没有广延时它所发射和吸收的虚光子可以有无限大的频率; 但是光子能量越大, 它的辐射与吸收对外场的依赖就越小, 所以当它在外场影响下被改变了的那部分电磁质量却是有限的, 总体上表现为有限的电子质量。

除了上述在 $k \rightarrow \infty$ 时出现的“紫外”无穷大之外, 在计算中还出现另一类无限大, 叫做“红外”发散, 因为它出现于 $k \rightarrow 0$ 的极限情况中。布拉赫 (Bloch) 等人发现: 这类发散准确地跟包含有一个多余的终态低能光子的康普顿散射项相抵消, 而且, 在 α 的级别上, 发生着这样准确的抵消。于是, 红外发散问题就解决了。

1940年, 泡利给出了自旋与统计法相关性的证明, 盖棺论定地对场的量子化提出了清晰的要求: 要求用对易子去使整数自旋场量子化, 而用反对易子去使半奇整数自旋场量子化; 否则, 将会导致矛盾。

后来, 施温格与朝永振一郎为消除二级修正项中的无限大提供了一个重整化程序的形式表述。与此同时, 费曼发展了一种便于计算的图形

技术。戴森证明了上述两种表述方式的等价性，同时也论证了重整化程序不仅对于二级项而且对于一切级别均有效。

第一部分附录

1. 聊聊氢原子

氢原子看起来非常像只有一颗行星的太阳系，尽管从量子力学的眼光来看，那唯一的一颗电子并不像地球绕着太阳转一样有着固定的轨道。

众所周知，氢原子核外的电子在原子核的电磁引力（而不是万有引力）作用下绕核运动。并且由于库仑力和万有引力形式上都是平方反比力，你很容易想象，也许电子就像行星一样绕着原子核作椭圆运动。“这是开普勒第一定律，只不过这个‘恒星’小了点罢了。”你也许会说。开普勒是17世纪最伟大的天文学家，他给出一个定律，说的是在平方反比引力的作用下，质点将作圆锥曲线运动——这个定律可以通过微积分在数学上给出证明。因此，当卢瑟福在20世纪初提出氢原子的“太阳系”模型之后，大家都觉得没有什么不妥。难道不是吗？太阳系这样庞大的结构，竟然和微乎其微的原子结构是一模一样的！我们在探索这个世界大尺度结构的同时，发现物质的基本结构和它如此相似，从哲学的角度来说这是合理的，所谓“一花一世界”，当然大家认为它是对的。

那个时候还没有量子力学，就连量子力学的主要构建者海森堡也还只是一个研究生罢了。好吧，大家就按照研究太阳系的方法研究氢原子，看看这最小的“太阳系”有着怎样秘密。

首先，核外电子当然是绕着原子核作椭圆运动，并且和开普勒第三定律说的一样，它转一圈所用时间的三次方正比于这个椭圆半长轴的平方。那么，这样的话电子应该作加速运动，按照麦克斯韦的电磁理论，它就会发射电磁波、损耗能量，进而落到更接近原子核的轨道上……如

此反复，根本就不可能存在稳定的原子啊！这里我们要想到一点，地球也在作加速运动，为什么地球就能不发射电磁波，从而日复一日，年复一年地绕着太阳一直转？那是因为地球所带的正电荷和负电荷一样多，整体呈电中性；可是电子阴阳失调，只带了一个负电荷，因此呈负电性，从而会发射电磁波。这样一来，人们发现氢原子并不像太阳系这么单纯，其背后一定隐藏着不为人知的秘密。

最先解密的人是丹麦物理学家玻尔，许多人认为他是爱因斯坦一生的对手：他们在同一年领取诺贝尔奖，却为了量子力学吵得不可开交。那时候的玻尔还很年轻，他恰好在卢瑟福的实验室工作。他奇迹般地想到，电子绕原子核运动的轨道可能不是连续的，这样就能解释氢原子光谱问题。实验中发现，氢原子的光谱并不是连续谱，而是一条条分立的谱线。这很容易理解：地球绕太阳转的轨道如果更近一点，它的能量就比原来要降低一点；如果地球轨道的半径可以连续变化，那么这个能量值也能是连续的数值。可是氢原子外边的电子却不是这样，分立的线状光谱表明原子的能量只能是分立的数值，从而说明电子的轨道半径也只能是某些特定的数值。玻尔提出这样的假设：①原子只能处于一系列不连续的能量状态中，在这些状态中原子是稳定的；②原子的定态是不连续的，因此电子的可能轨道的分布也是不连续的。这样的理论无疑在轰击经典物理学的大厦，而且它和人们的直觉是格格不入的。试想一下，它相当于说存在一根弹簧，它只能拉出10厘米、20厘米的伸长度，而不能拉出15厘米的伸长度——但是从直观的角度来看，我们必须先拉长到15厘米才能进而到20厘米！

其实我们还是可以接受这些“不连续”的概念的。设想一下，在一根琴弦上振动的波，琴弦的长度一定是半波长的整数倍；在电脑里传输的数据大小，也一定是1Kb的整数倍。甚至我们可以想象，现在流通的最小面值纸币是一角钱，如果你遇上一个乞丐并且善心大发想给予施舍，你一定也只能给一角钱的整数倍。如果你把一角钱撕成两半，然后给乞

丐其中之一，那么乞丐一定不会鸟你——在这里，钱就是能量；那个乞丐，就像那颗只接受特定能量值的电子。

玻尔看着自己的两个假设，惊呆了。但是由玻尔假设可以很容易推导出光谱学中很重要的巴尔末公式，它精确描述了氢原子的一族谱线规律。于是玻尔得到大家的赞许，被邀请到各个地方演讲他的新理论。和现在的讲座一样，一百个听众里只有一个听众是认真去思考的，只有一个人敢和演讲者分庭抗礼。而玻尔在哥本哈根就遇上了这么一个愣头儿青——凡尔纳·海森堡。后者当时只是个二十出头的年轻人，跟着他的老师来打打酱油，却在观众提问环节问了玻尔一个很难回答的问题：“伟大的玻尔教授，请问电子在两个轨道之间跃迁需要的时间是多少呢？”玻尔被难住了，尴尬地下不来台，但是玻尔毕竟是玻尔，一个有经验的演讲者，他如是回答：“这是个好问题，年轻人。”现今很多教授都学会了用这招来逃避尖锐的提问，但是他们却没有学会玻尔的谦虚。在会后，玻尔邀请海森堡一起散步，允许他进一步阐述他的想法和所提的问题。

海森堡是幸运的，他想到一个更基本的问题。

根据一些文献的记载，海森堡称那次散步使他“开始了科学生涯”。一个人的成功，和时代背景分不开。而海森堡和一个完美的时代邂逅了——开尔文勋爵所说的经典物理学的大厦上空的两朵乌云电闪雷鸣，经过这几年的发展，这座大厦成为危楼，即将坍塌；所有人都在等待着救世主的降临。在这个时候，海森堡遇到了索末菲，认识了玻尔，而且对量子论发生了兴趣。因而，这次散步，从科学史的角度来说，也开启了一个时代！

那一年海森堡才20岁，许多中国的年轻人20岁的时候还在读本科二年级，但是海森堡却已经是索末菲的研究生了。年轻者无畏，海森堡很有自己的想法，他思考着玻尔的量子论，试图用更加基本的理论取代它。“当电子从一条轨道跳跃到另一条轨道的时候它在做什么？”海森堡

思索着，“怎么知道电子一定作圆或椭圆运动呢？”在前文我们说过，电子受到原子核的库仑力作用类似于地球受到太阳的万有引力，因而根据牛顿力学定律我们马上就可以知道电子做的只能是三类圆锥曲线运动。当电子没有脱离原子核的束缚，那就应该是作圆或者椭圆的运动了，进而我们才能讨论电子的轨道。“可是，”海森堡想，“如果换一个角度去看，会怎样呢？”海森堡指的是与牛顿力学体系齐驱并驾的另一个经典力学体系——分析力学。

分析力学很难说不是数学的一个分支，建立它的两个重要人物拉格朗日和哈密顿都是数学家。拉格朗日更是说他在试图把力学变为数学分析的一个分支上建立了这门学科。因而，分析力学比牛顿力学更玄一些，对于那许多左手就能写出牛顿方程的聪明少年而言，这真是引人入胜。在分析力学的哈密顿形式里，把描述物体运动的物理量扩展为两个，即广义动量和广义坐标，从而把一个二阶微分方程（牛顿第二定律的方程就是一个二阶微分方程）降为两个一阶微分方程。在它的体系里，认为坐标和动量是平权的，是一对共轭的正则变量。因而我们再来描述体系的运动状态，是在横坐标表示广义坐标，纵坐标表示广义动量的相空间内用一个点来表示——这有点类似于大家学过的s-t图或者是v-t图，但是这个更为深刻一些。

总之，我们现在从分析力学的角度去看氢原子电子的运动，我们有一对正则变量：电子的坐标，电子的动量。哦，不对！我们能不能同时拥有电子的坐标和电子的动量？这在经典的世界看来是可以的。但是，在量子力学里，这恰恰是区别于经典的关键。试想像电子这样小的东西，还在运动；当我们测量它的坐标的时候就会给它无法避免地施加一个影响，这样就会改变它的动量；当我们测量它的动量的时候，同样会给它施加一个影响，就会改变它此刻的位置。因此，我们根本不可能同时拥有电子坐标和动量的准确值！

哦，这个想法真是太大胆了。如果按照这样的观点，电子就不会有

所谓的“轨道”，汽车也没有办法避免逆行，每个人也没有办法把自己摆在正确的位置上，世界乱了套。就在我们一筹莫展的时候，一个常数 h 现身人间拯救世界。

h 是普朗克常数，它是一扇分隔经典世界和量子世界的大门。海森堡发现了量子世界的一个公式

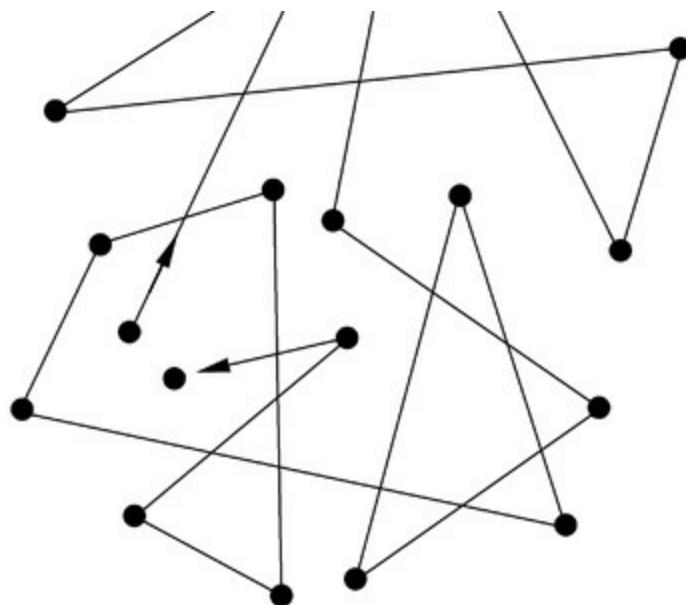
$$[x, p] = xp - px = \frac{ih}{2\pi}$$

此公式称为基本的对易关系。根据量子力学原理，可以推导出不确定关系

$$\Delta x \cdot \Delta p \geq \frac{h}{4\pi}$$

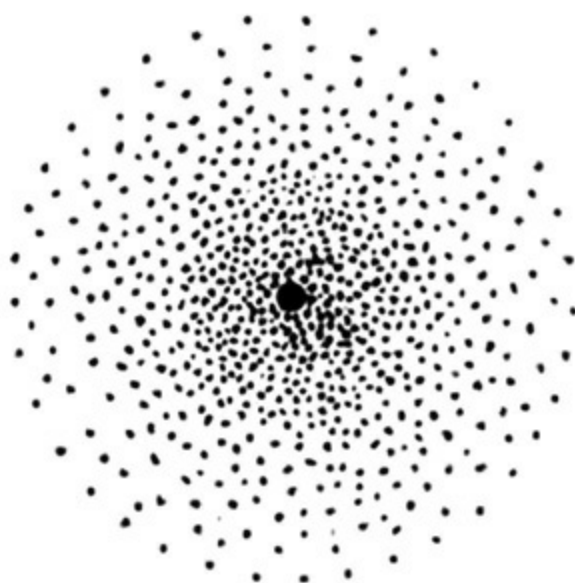
这第一个式子我们看到氢原子坐标和它的动量的乘积得到的东西跟它俩的顺序有关系，第二个式子可以看出坐标的变化量和动量的变化量不能同时为零。如果量子世界满足这样的数学，那么上文的结论“我们根本不可能同时拥有电子坐标和动量的准确值”就是正确的。但是这样也不会引起车辆逆行、世界混乱的结果，因为普朗克常数实在是太小了，它的值是 $6.6260693 \times 10^{-34} \text{J}\cdot\text{s}$ ，如果我们计算一颗子弹的不确定度，那就小得多；但是如果我们计算氢原子电子的不确定度，那就是可以和氢原子电子大小相比拟的数值了。因此，普朗克常数的出现，使得人间重现光明与秩序。

实际上，使得我们更加愿意相信电子是没有按照一定轨道运动的理由，是我们从来没有在实验中观测得到电子的轨道！我们总是在这一刻发现电子大概在这个区域，下一刻却跑到了另一个区域，局域地看，它似乎是在作杂乱无章的运动。



布朗运动

但是如果叠加了大量的数据，我们就能发现它出现在某些区域的概率特别大，而在某些区域出现的概率几乎是零。



（注意图上只有一个电子，那么多的点实际上是每次电子位置的叠加）

因此，海森堡认为我们必须放弃“电子轨道”的概念。他像“王”一样地宣布：“从今往后，我们只相信实验观测到的物理量。而对于那些臆

想出来的东西，我们一律摒弃！”整个物理学江湖为之震撼！一些江湖大佬纷纷质问他，什么是需要丢弃的呢？是速度、是加速度还是力？海森堡登上主席台，神色坚定地说：“经典物理里可观测量在量子力学实验中测得的都是平均值！”海森堡进一步阐述他的想法：“量子力学里，力学量要用算符来代替。因此，电子只是概率地出现在氢原子周围，并没有确定的轨道。”海森堡接着用它的理论做了一些计算，完美地解释了一些现象，得到大家的信服。

但是，就连海森堡本人也不知道他“创立”的量子力学所用的数学是什么，怎么会有 $xp - px \neq 0$ 呢？如果这是两个数的话，无疑是令人难以接受的。现在上过大学本科学理工的同学都能知道，如果 x 和 p 是两个矩阵的话，它们之间有这种关系真是太正常了。但是，那个年代的物理学家，学过线性代数的寥寥无几，所以作为一流的物理学家，海森堡不知道这就是矩阵的乘法也不足为奇。而且更令他骄傲的是：“我虽然没有学过矩阵，但是我却可以创造我需要的矩阵！”物理学家的神奇，有一点在于他可以天马行空，可以发明和创造自己需要的一切武器，比如牛顿发明微积分。直到后来海森堡拿着他的理论去见他的一个长辈——也是很牛的物理学家波恩——波恩见多识广，告诉他这个东西叫做矩阵，你去看线性代数的书吧。后来海森堡靠着这一数学手段，成功开辟了量子力学的一条途径，矩阵力学。

但是，任何一个本科学量子力学的中国学生，他更为熟知解决氢原子问题的方法不是来自矩阵力学（我们管他叫做代数解法），而是从薛定谔方程出发，解这个微分方程。薛定谔的这条通往量子力学的道路，我们称为波动力学。

其实，薛定谔的出发点也非常简单，甚至比海森堡要容易一些。 $xp \neq px$ ，那么如果你不知道矩阵的乘法具有这种性质，那么即使你是一个小学生，你也能想到如果 x 表示+号， p 表示×的话，那么先加后乘和先乘后加之间一般也是会有差别的。薛定谔的波动力学就是这个样

子，只不过他更深远一些，他认为像坐标、动量这些力学量在量子力学里可以是算符，比如动量就是偏导数算符（坐标表象）。这个时候他看到一个年轻的博士生写的论文，讲的是不仅电子具有波粒二象性，任何物质都有波粒二象性，只不过由于普朗克常数的限制，大物体的波动性不易表露出来。于是薛定谔信手拈来，在阿尔卑斯山的落日余晖中写下了人类历史中非常重要的一个方程——薛定谔方程

$$\left(-\frac{h^2}{2m} \Delta^2 + V \right) \varphi = i h \frac{\partial \varphi}{\partial t}$$

这里 φ 叫做波函数，他一般是关于位置和时间的函数，它模的平方就是在某时刻某个位置测得粒子的概率。如果没有学过高等数学，我们看着这个方程有些头疼，可能隐约会觉得美丽，更多是觉得一头雾水。其实这个方程并不困难，不管是从形式上来看还是从作用上来说，它都只是量子力学的牛顿第二定律 $F=ma$ 罢了。 $-\frac{h^2}{2m} \Delta^2$ 这一项代表粒子的动

能， V 则是粒子所处的势能。总之，不管怎样，薛定谔写出了这个方程，并没有证明它（在量子力学里，这是一个假设，不能被证明），但是根据这个公式计算的一些结果，却与实验数据吻合。其中就包括氢原子。

2. 海森堡的墓碑



3. 尾声：测量光谱和波恩的彩虹

总的来说，量子力学波函数的概率解释归功于波恩，他用彩虹把量子世界和经典现实世界连接起来。

(1)

夏天的雷雨后，或许可以看到天空中挂着一道彩虹。旖旎的山水风光和天际的彩虹，似乎在告诉人们，生命是多么的短暂，一生都应该像彩虹一样绚烂。彩虹一头连接着天穹，一头弯向大地。

在量子力学和现实世界之间，同样需要一道彩虹，这道彩虹就是波恩（Born）。我们可以再谈一点测量光谱的问题，来看看测量为什么那么重要。

这个世界，是非常模糊的，如果不仔细思考，不仔细分辨，直观主义会带来足够多的误会。梁实秋曾经问过“有一件事情，男人站着做，女人坐着做，小狗跷着一条腿做”。这是一件什么事情？梁实秋说，这件事情是握手——这个答案是具有技术性的。

同样道理，当物理学家去测量一束光的强度的时候，需要更加高的

技术性。在本书第10章，读者们已经了解到，如何测量光的波长。这里则要谈论另外一个问题，那就是如何检测一个光源随波长的光强分布（光谱）。

首先，有一个基本的问题，那就是有没有一个绝对标准的光源。

家庭用的钨灯，是一种热辐射发光的灯泡，这种灯泡发出的光谱，肯定不是严格的黑体辐射谱，因为钨会产生一些比较尖锐的特征峰。

那么有没有标准光源呢？

这个问题也许你觉得一点也不重要，那么，我们可以问性质一样的另外一个问题：“如果一个姑娘穿的是红衣服，你觉得这衣服真的是红色的吗？”

理论上存在的绝对标准的光源，大约只有两种，一种是黑体辐射，基于普朗克等人的计算。另外一种，则是同步辐射，基于施温格的计算。这两种连续光谱是相对比较纯洁的光谱。也被认为是，真正标准的光谱。

一般做遥感研究的人，研究卫星探测到的植被对太阳光的反射光。

其他行业也是一样，探月卫星上安装有X射线探测仪等光谱仪器，其主要功能就是发射一束入射光，然后研究反射光。

在这个意义上，入射光必须是已知的，最好是标准光源。——这就需要对光源发出的光谱有清晰的认识。

当一个光源发光以后，比如说一个氙灯发光以后，有一个光谱，这需要仪器去测量出来。除了单色仪可以扫描波长以外，需要一个检测光强度的仪器。探测光强度的装置有很多种类，比如数码相机用的CCD，或者说一个光电倍增管。

以光电倍增管为例。当不同波长的光照上去以后，它的响应曲线并不是一条直线，而是一条曲线（依赖于量子效率）。因此，你能够测量到的光谱，其实是入射光谱和光电倍增管响应曲线相乘以后的乘积曲线。

因此，当我们真的是探测一个光源光谱的时候，我们在仪器上能看到的光谱图，实际上并不是真实的，而是携带了仪器的信息。

可是，真实的光谱到底应该是什么样子的呢？

这只是没有读过量子力学的人才会问的傻傻的问题。本书读者应该很清楚了，真实的光谱是不存在的，这个世界是基于观测的，而观测是仪器依赖的。

从某种意义上说，我们是非常幸运的，因为我们生活在一个因为量子效应而带来的各种模糊的世界里，所以才会如此精彩。而关于量子力学的波函数在观测上的意义，其概率解释，则归功于波恩。

(2)

20世纪30年代钱德拉塞卡虽然得到了关于白矮星的理论，但他因为受到爱丁顿的严重伤害差点离开了天体物理和相对论，转而研究天空为什么是蓝色的。但从他开始，人们真正开始关心恒星的整个命运，并且猜想在恒星质量很大的时候，黑洞将应运而生。中子星是当时物理学的一个十分交叉的领域，在1932年之前，还没有中子的概念，但一旦中子被确实，中子星的概念马上应运而生（正如夸克被确实，夸克星的概念也浮出水平）。当恒星演化成为黑洞的过程中，星体内部的电子和质子被压在一起，这两种具有相反电荷的粒子被迫走在一起，形成了不带电的中子，因为电子和质子的自旋角动量都是 $1/2$ ，根据量子力学自旋角动量的耦合，它们的结合不可能只得到自旋为 $1/2$ 的粒子，而中子的自旋为 $1/2$ ，所以在电子和质子被“逼婚”的过程中，中微子必须介入这桩“婚姻”，中微子的自旋也是 $1/2$ 。要预言中子星在什么质量区间内存在，其中的计算非常物理，需要估计中子气的状态方程，理想气体的状态方程，是很多人熟悉的，就是所谓的克拉伯龙方程 $pV=NRT$ 。但中子气的状态方程，是什么样子，很少有人能写出来。1939年奥本海默和

沃尔科夫通过计算建立了第一个中子星的模型，他们的计算也只能给出一个比较粗放的结论，大致是说，如果恒星的质量是2~3倍太阳质量的话，那么，恒星最后会演化成为中子星。

奥本海默是美国的原子弹之父，他的故事富有传奇的色彩。奥本海默是量子力学奠基人之一——德国物理学家玻恩的学生，也正是奥本海默等人把欧洲的理论物理搬到了美国。因为中子星的密度极大，不能使用牛顿的万有引力，又必须要用到广义相对论。真正能做这些估计和计算的，在当时的人群里已经寥寥无几了，研究者不但要懂得广义相对论，还必须及时地跟上量子力学发展的步伐。

这一切与量子力学有了千丝万缕的联系。从发现电子，发现中子，这些历史过程之中，人们渐渐地完善了量子理论，逐渐清晰地描绘了微观世界的景象。

(3)

爱因斯坦搞出一个狭义相对论的公式，能量等于质量。

奥本海默是原子弹之父。他是美国人，出生在有钱人家，他在欧洲跟玻恩做研究生的时候，听报告总坐在前排，因为他已经习惯于在人家作报告时冲上去说，照我看这里用这个方程会比较简单，然后拿起粉笔在黑板上狂写。很多教授作报告，只要奥本海默在场，都是报告作着作着就渐渐沦为一个配角。奥本海默年轻的时候似乎有点从来不顾及他人之感受。但他的确是少数的几个实干家之一，学会了近似处理问题。在量子力学里就有玻恩和奥本海默近似研究的是一群分子的群体行为，非常之物理。等他后来回美国领导原子弹研发，在洛斯阿拉莫斯，集中了大量研究人员，曼哈顿工程细致缜密而且庞大过人，奥本海默体现出杰出的领导才能。原子弹爆炸成功后，奥本海默成为名人。正当他春风得意，开始有人说他是共产党，有人说他是苏联的间谍，有人说他

是.....，反正是积毁销骨，他付出了名人的代价。原子能委员会或者FBI对他开始了不断地深入持久的调查，剥夺了他研究氢弹的权利。

奥本海默，开始了他漫长的像岳飞在风波亭里那样的人生历程。

研究氢弹的任务交给了泰勒，泰勒在对奥本海默的调查中做了对后者不利的证言。泰勒先计算了氢弹的威力，发现一旦氢弹爆炸，整个大气层就要燃烧殆尽。这就是泰勒，一个很有想法的人，杨振宁先生说过，泰勒是一个想法很多的人，他的想法，90%是错的，但他很敢想，是一个不错的物理学家。奥本海默后来的人生是一幕悲剧，他的人生经历也许说明政治的残忍。回头让我们来看看其导师波恩的人生，后者的人生就是架在神秘的量子力学世界和人类的经典世界的桥梁之上。

玻恩是一个犹太人，他先在德国柏林大学做教授。1921年，他还接替德拜成为哥廷根大学物理系主任——当时，海森堡从慕尼黑大学博士毕业以后就来哥廷根大学投奔玻恩，并且在这期间花粉过敏，去休假的时候发现了量子力学的矩阵形式。1926年，薛定谔写出量子力学波动方程后，薛定谔本身对这个波函数的物理解释存在一定的困惑，后来玻恩给出了波函数的概率解释——好像一道色彩绚烂的彩虹，一端伸向天空，一端拉起大地。薛定谔的量子力学波函数在空间每点指定一个复数 Z ，对于这个复数 Z ，物理意义完全不清楚，玻恩用高中数学的手法取了这个复数 Z 的模，他把模的平方（实数）解释为粒子出现在该空间点的概率。正是这貌不惊人的一步把神秘的量子力学和经典的测量联系了起来。对 Z 取它的模平方这在复变函数来说不是一个全纯

（holomorphic）的变换，换句话说，把量子信息转化为经典信息的时候，玻恩的概率解释破坏了量子理论的全纯性。全纯函数就是复数形式的解析函数，它定义在复平面上。波函数的概率解释也引出了量子力学的测量问题：测量引起波函数的塌缩。

以上的说法已经是把当时玻恩的工作高度数学化了，玻恩当时是从微观粒子相互碰撞的散射过程中，渐渐地摸索出这个概率解释的。

1933年希特勒在德国掌权后，玻恩由于犹太血统关系被剥夺了教授职位和财产。他流亡到英国，他逃到英伦的时候风声很紧张，他一下轮船就看见路对面打着一横幅，上面赫然写着：Born to be hanged。玻恩大吃一惊，心里大喊，莫非，莫非，真是天亡我也，难道在我坐船来英国的途中英国已经被希特勒占领？

玻恩后来才想到的是自己的名字Born在英文里有“出生”的意思。1954年他和黄昆合著的《晶格动力学》一书，被誉为固体物理理论的经典著作。玻恩还和人合作，写了一本《光学原理》，这本书成了光学领域的经典著作。

现在在哥廷根，玻恩的墓碑上刻着量子力学中最重要的不对易关系式：

$$pq - qp = h/2m$$

在墓碑上刻数学公式，一般人是很少用的，玻耳兹曼也用过，他用的是 $S = k \ln W$ 。这些人全是有点天才气质的，尤其是玻耳兹曼，他是统计物理的杰出领袖——薛定谔刚去维也纳上大学的时候，玻耳兹曼刚好自杀去世，因此，年轻的薛定谔没有能见到他当年的偶像——当然后来薛定谔自己也成了偶像。

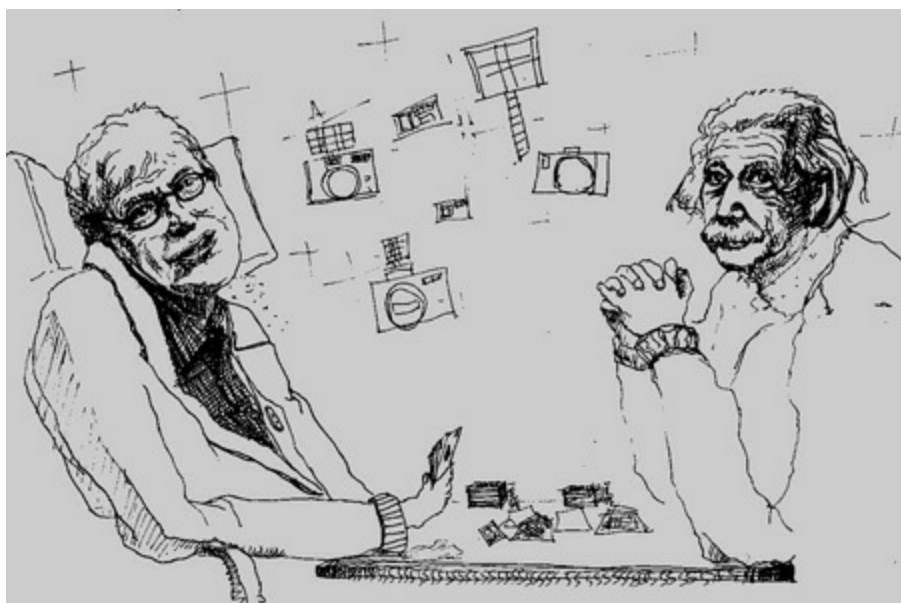
玻恩的墓碑上的不对易关系其实就是“海森堡不对易关系”，这是量子力学数学基础之一， p 和 q 本来表示粒子的位置和动量，是实数，但玻恩墓碑上它们全是算符。算符是作用在希尔伯特空间上的，因为希尔伯特空间里可以找到一组基底，所以算符可以被认为就是矩阵。这个关系是量子物理区别于经典物理的关键之处。（当然狄拉克后来认为，量子物理区别于经典物理的关键之处不是不对易关系，应该是量子物理的波函数的相位不定性。在两个波函数差一个相因子，在物理上无法区别，它们代表同样的量子态。）“海森堡不对易关系”原理也叫不确定原理。不确定原理看上去是如此简单，以至于几乎没有人能完全明白——虽然数学家可以从傅里叶变换中天然地得到它，但在物理上真是很费解的。

因为这个不确定原理，使得量子力学几乎和爱情一样微妙。谈恋爱的男女，女生往往会提出这样的问题：“你确定爱我吗？”或者“毕业了你会不会离开我？”这个时候，男生往往不能做出很确定的回答。在爱情生活上的不确定性，描绘普遍的人性，流行歌曲歌声此起彼伏：“随缘分过去你不再问 / 不懂珍惜此际 / 每每看着我伤心 / 只因你看惯我的泪痕 / 对你再不震撼 / 看见了都不痛心 / 如何像戏里说的对白 / 想恋一生一世 / 说了当没有发生 / 思想已永远退不回头 / 爱过痛苦一生 / 沾满心中的泪印。”

第二部分 相对论

SUNRISE

A story of the Quantum theory
and Relativity



爱因斯坦和霍金玩扑克 绘画：涂明

22 乡下的月光

(1)

1666年，一个神情憔悴的年轻人在夕阳下站在河边，英伦的伍尔索普一带，宿草盈阡，这个年轻人看着流水平桥，天边又有一群乌鸦飞过，带走无尽凄凉。瘟疫似乎没有尽头，此地孤寂而冷清，这个年轻人就这样一直孤零零地站着。

附近的奶牛场里，一个姑娘正蹲在地上抚摩着奶牛的丰硕的乳房。这个姑娘身形曼妙，一笑一颦皆令人春心荡漾。

牛顿对这个姑娘，竟是颇有好感。但又不知怎么的，竟又想起自己的身世。他是一个遗腹子，20年来人世飘零，1642年的圣诞节前夜，在英格兰林肯郡沃尔斯索浦的一个农民家庭里，牛顿诞生了。牛顿是一个早产儿。接生婆都担心他能否活下来。谁也没有料到这个看起来微不足道的小东西最后会成为一位震古烁今的科学巨人，并且活到85岁的高龄。牛顿出生前三个月父亲便去世了。在他两岁时，母亲改嫁。从此牛顿便由外祖母抚养。11岁时，牛顿的继父去世，牛顿才回到了二次寡居的母亲身边。大约从5岁开始，牛顿被送到公立学校读书，12岁时进入中学。少年时的牛顿资质相当平庸，成绩一般。他绝不是陶哲轩这类的少年英才，但牛顿喜欢读书，喜欢看一些介绍各种简单机械模型制作方法的读物，并从中受到启发，自己动手制作些奇奇怪怪的小玩意，如风车、木钟、折叠式提灯等。药剂师的房子附近正建造风车，小牛顿把风车的机械原理摸透后，自己也制造了一架小风车。推动他的风车转动的，不是风，而是老鼠。牛顿将老鼠绑在一架有轮子的踏车上，然后在轮子的前面放上一粒玉米，刚好那地方是老鼠可望而不可即的位置。老鼠想吃玉米，就不断地跑动，于是轮子不停地转动。他还制造了一个小

水钟。每天早晨，小水钟会自动滴水到他的脸上，催他起床。但快乐的日子很短暂，后来，迫于生活，母亲让牛顿辍学在家务农。但牛顿对务农并不感兴趣，一有机会便埋首书卷。每次，母亲叫他同她的佣人一道上市场，熟悉做交易的生意经时，他便恳求佣人一个人上街，自己则躲在树丛后看书。有一次，牛顿的舅父起了疑心，就跟踪牛顿上市镇去，他发现他的外甥伸着腿，躺在草地上，正在聚精会神地钻研一个数学问题。牛顿的好学精神感动了舅父，于是舅父劝服了母亲让牛顿复学。牛顿又重新回到了学校，如饥似渴地汲取着书本上的营养。这一路走来，受尽了世态炎凉，牛顿也变得沉默，他得了少儿爆怒症，经常想打人骂人。后来他终于从剑桥大学毕业，却又赶上了瘟疫，于是只好逃到乡下来避难。人生黯淡。牛顿此刻想到，开普勒的行星运动三定律，背后似乎暗藏玄机。开普勒是一位天文学家，他的行星运动三定律都是在平面几何的定理。

定律如此说：

1. 行星运动的轨迹是平面上的一个椭圆，太阳在椭圆的一个焦点。
2. 行星矢径在单位时间内扫过的面积相等。
3. 行星运动一周的时间的平方和半径的三次方的比率是一个常数。

这三个定律，真是有点奥妙，牛顿心想，晚上之前，他决定不再想那个牛奶场里的姑娘，好好地把开普勒的东西捋一捋。

就在这天马行空的念想之中，落日熔金，天色也就渐渐地黑了。牛顿蹲在草地上，开始找了一根树枝，在地上画起来，他在前几天发明了一种叫微积分的数学工具，帮了他的大忙，他越写越觉得事情是如此这般的，也许开普勒背后确实有神秘的力量。

月亮开始升起来，奶牛场的姑娘已经回家了。

牛顿抬头望着月光，狡黠的月亮像是大饼一样挂在夜幕之上，牛顿

觉得饥肠辘辘。再回头，看见远处旷野上突兀地立着一棵苹果树，牛顿大步流星地朝那棵苹果树扑去。

空气是那么迟滞，静穆的夜色，把乡下的这一切掩映得如在一层看不透的浓雾之中。牛顿刚跑到苹果树下，还没有站稳，就被一个从树上掉落的苹果砸在鼻梁之上。

“哎哟”，牛顿好像被鬼掐了，大叫一声，低头看那砸人的苹果已经滚落在一旁的草丛中。

牛顿愣住了，过了半天，才一屁股坐在苹果树下，慢慢地捡起方才那个落地的苹果，用手擦了一下，慢慢地啃了起来。

天空中依然是月光惨淡，冷漠地不知道如何收场，月亮似被冰冻，一动不动。牛顿吃完苹果，又抬头看了一眼月亮，觉得月光似乎温暖了一点儿，开始在天空中移动。他似有顿悟，猜测苹果落地，和月亮在天空中公转，似受到同一个力量的牵引。

这个神秘的力量，对世间万物都是一样的，于是，他称为万有引力。

(2)

万有引力的大小与距离 r 的平方成反比，这是因为空间是三维的，只有这样，引力场的高斯封闭面积分才可以得到引力荷。如果空间是二维的，那么，同样的积分将给出引力场的大小将是距离 r 的倒数——从这个意义上说，三维空间的结构与引力场的强度本身就处于一种相互调节的状态。

牛顿是一个工作很努力的人，他创建了一个力学体系，包罗万象。比如，他的牛顿第二定律是一个对时间的二阶微分方程，这个方程依赖于惯性坐标，但到底什么是惯性坐标？牛顿说，满足这牛顿第二定律的坐标就是惯性坐标。这样一下就把大家都绕进去了，因此，牛顿把这个

世界打扮了一番，变成另外一个模糊不清模样。牛顿理论还有更加隐秘的硬伤，那就是他的引力理论只能计算质点之间的相互吸引，质点是无限小的点，自然是不存在的，所以，如果要计算两块石头之间的万有引力，大家被迫要做一些很复杂的积分，这些积分很容易把很多人吓死，因为实在是复杂得可以。所以，如果再要求这两块石头旋转起来，再计算二者之间的相互万有引力是多少，这几乎成了不可能完成的任务。同时牛顿的万有引力是平坦背景空间上的瞬时超距作用，也就是说，假如现在太阳突然消失了，那么地球马上就会因为惯性甩出去——这是离心现象。假如按照牛顿万有引力，旋转星球所辐射引力波的星体会越辐射质量越大——这情境很像是一个在健身房汗流浹背的女孩子，人们看到她流了很多汗，但身子却同时在变胖。所以，牛顿万有引力是无法处理引力辐射的。

牛顿在月光之中发现的万有引力结构，实用但不优美。他的那个和苹果有关的故事，激励了一代又一代人，甚至也鼓励到美国苹果公司的乔布斯。

但在那个时代，牛顿作为科学家的杰出代表就像在中国企业家当选人大代表一样。他于1689年当选为国会中的大学代表。作为国会议员，牛顿逐渐开始疏远给他带来巨大成就的科学。他花了大量的时间在和同时代的著名科学家如胡克、莱布尼兹等进行科学优先权的争论上，很多在媒体上以牛顿的朋友名义发表的支持牛顿的文章都出自牛顿本人之手。又过了20年，牛顿已经是皇家造币厂的厂长，他推动了金本位与银本位的挂钩，而且参与了股票市场的运作。1720年的英国股市同2007年的中国股市一样，是一个大牛市。其中一只股票格外出众，是大牛股中的佼佼者——这就是有着国家垄断权利的南海公司。1720年1月南海公司上市，股价128英镑，之后在股民的追捧下一路高涨猛进如当初的中石油。4月，作为造币厂厂长的牛顿忍不住了，投了7000英镑购买南海股票。股票继续高涨，仅仅两个月，当谨慎的牛顿把股票卖掉后，竟然

赚了7000英镑！可是刚刚卖掉股票，牛顿就踏空了。到了7月，股票价格竟然达到1000英镑，增值近8倍。像所有踏了大空的股一样，牛顿后悔极了！

然而牛顿不服输，经过他严密的科学计算，牛顿又决定买入南海股票，并且还借钱加大投资。结果股票开始一路狂跌，到了12月又跌回128英镑。牛顿没有来得及脱身，亏了2万英镑——他10年的工资！当然牛顿不愧是大科学家，输光了钱后，他明白了股市，他坦白地说：“我能算出天体运行的轨迹，但算不出人性的幽暗。”

晚年的牛顿在伦敦过着富丽堂皇的生活，他被安妮女王封为贵族。此时的牛顿非常富有，被普遍认为是生存着的最伟大的科学家。他担任英国皇家学会会长，在他任职的24年时间里，他以铁拳统治着学会。没有他的同意，任何人都不能被选举。

1727年3月20日，伟大的艾萨克·牛顿逝世。同其他很多杰出的英国人一样，他被埋葬在了威斯敏斯特教堂。

牛顿曾经来过，他最重要的发现，是在乡下的那一年，看到的月光的秘密。

外一篇 盲剑客

数学家欧拉，年轻的时候并不是一个瞎子。

他对刚体的问题，有一些研究。

一个刚体，比如一个太空中的陨石，如果它不受到任何外力的作用，那么它将如何运动呢？

这是欧拉关心的问题，这被称为自由刚体的运动。（欧拉本质上是一个数学家，在数论上发现了著名的恒等式，就是把黎曼级数和素数的无限乘积联系起来了。）

刚体的自由运动显然是有6个自由度的——确定原点的坐标需要3个

数，确定3个转动的角度方向又需要另外3个数字。在每一个瞬间，可以给这个既在平动又在转动的石头拍照，总可以得到它的质心的位置是3个坐标，然后它身上固定的一个三脚架有一个特定的方向，这样它又有3个转动的自由度。所以，一共是6个自由度。但是，因为自由不受到外力的质心是作匀速直线运动，因此，这个自由运动需要研究的部分其实可以看成是绕质心的旋转运动。

换句数学的语言，刚体的自由运动就是，绕质心的旋转运动，这是一个三维空间的三脚架的绕一个固定点的转动，这个旋转运动发生的位形空间是一个三维的流形，这个流形其实就是数学家们所说的李群 $SO(3)$ 。 $SO(3)$ 上的每一点，都表示刚体转动后的一个位置。在这个意义上，自由刚体的运动状态（位置和速度）可以用 $SO(3)$ 的切丛 $TSO(3)$ 来刻画——当然你可以考虑怎么量子化这个问题（就是第2章监狱里的群论讲过的， $TSO(3)$ 是这个李群的一个切丛。打个比方， $SO(3)$ 是一个篮球， $TSO(3)$ 就是让篮球在操场上滚一圈）。

这个 $TSO(3)$ 是一个6维流形，它之上的每一个点，就对应绕固定点的旋转的刚体运动一个特定的位置和它的速度。这个时候，如果没有守恒量，那么运动就发生在整个6维空间里，变得很可怕。幸亏有3个角动量分量和总能量守恒，所以可以去掉4个自由度，运动实际上只发生在一个 $6-4=2$ 维的子流形 M^2 里。在整个自由刚体运动的过程中，运动其实只发生在 M^2 上，而二维的子流形性情非常温和，彼此之间几乎没有什么区别，一般都是带几个洞 M^2 的环面，也就是只有亏格数的不同而已——因为没有其他额外的守恒量了，所以这个运动发生在曲面上，是不能唯一确定出一条一维的运动轨迹，这种状况被称为“不可积”的。欧拉在某一个时刻得到了描述绕固定点的旋转的自由刚体运动的运动方程，这个运动方程被称为“欧拉方程”。

很明显，刚体没有受到外力，所以它的角动量应该是守恒的。

角动量又是什么呢？

角动量和刚体的质量几何有关系，刻画质量几何的是一个张量——转动惯量——静止刚体对刚体上某一点的质量几何可以用一个椭球面来刻画。

角动量正是转动惯量矩阵和角速度矢量的积。因此，如果角速度矢量方向不是转动惯量矩阵的特征矢量方向，从基本的线性代数可以知道，角动量矢量的方向可以与角速度矢量方向有一个夹角。

因此，很明显，虽然角动量是一个守恒量，它与时间无关，但转动惯量和角速度却可能与时间有关。人们可以从转动惯量中构造出四极矩来，如果这个四极矩随时间变化，那么在爱因斯坦引力中，它将成为引力波辐射的源。而这是牛顿力学所无法处理的理论硬伤。

23 牛顿引力的高级版本：拉普拉斯方程

(1)

帮国王拿破仑打工的学者，不但有钻石王老五傅里叶，还有另外一个数学家拉普拉斯。国王拿破仑是一个数学爱好者，他曾经有一个拿破仑定理，是很有意思的。定理说，任何一个三角形，各边上各作等边三角形，接下来将这三个三角形的重心联结起来，那么就必定是一个等边三角形。据说这个定理是一个叫拉普拉斯的数学家发现以后赠送给国王拿破仑的。其实与这个拿破仑三角形类似的还有一个叫做莫雷三角形的定理：在任意三角形的三个角做三等分线，就近相交得到一个小三角形，这个小三角形一定是正三角形。此莫雷三角形定理博大精深，作者见过的最精妙的证明来自数学家康内斯（Connes），他曾经用高屋建瓴的群论方式得到了一个证明。

话说回到法国大革命的时代。数学家拉普拉斯的数学才能，肯定是远过于拿破仑。拉普拉斯研究的东西很广泛，他甚至还研究过我们头顶的星空。尘世间有所谓康德-拉普拉斯星云说，意思是说，我们地球上的一切，最早都来自星云的演化。拉普拉斯问的问题也很简单：我们的太阳系是稳定的吗？牛顿认为是稳定的，因为“上帝会在合适的时间加以调节”。但拉普拉斯用了25年写了五卷《天体力学》，证明了一大堆关于扰动、轨道之类的结论，然而，还是没有完全搞明白这个问题，但他大致可以证明，我们的太阳系是稳定的。在他在书里他偷偷引用了拉格朗日、勒让德等人的工作，因为他没有提及原作者的贡献，所以，他被视为一个势利小人。

拿破仑觉得拉普拉斯很不错就让他从政。

拿破仑问道：“在你那个关于太阳系稳定的伟大的证明中上帝扮演

了什么样的角色？”

“陛下，我不需要这个假设。”拉普拉斯掷地有声地说，“大自然的全部结果不过是少数几个永恒定律的数学推论。”

太阳系的稳定问题，其背景只是牛顿万有引力。牛顿万有引力，在空间的每个点都有一个力的大小和方向，因此，我们说，力是一个矢量场，但我们也可以把矢量场看成是一个标量函数的微分。这个标量函数，就是势能——这就好像对于台风，我们可以研究其风速，也可以研究其气压。而在真空中，引力势能满足拉普拉斯方程（这个本质上就是说，在真空中，引力场的散度是零，或者等价地说，就是真空中没有质量分布）。

拉普拉斯方程有一个非常好的性质，就是你随便画一个球面，那么引力势在球面上的平均值等于球心处的引力势。据此可以推论，边界上才可能出现势函数的最大最小值。这个神奇的兴趣就好比说，北京的气温，一定等于北京周边的4个城市的气温的平均值。如果某一天保定的气温是 20°C ，天津的气温是 23°C ，张家口的气温是 18°C ，唐山的气温是 19°C 。那么北京的气温作为它们的平均数值就是 20°C 。

作者张轩中在工作的过程中，经常与拉普拉斯方程打交道，经常要研究质谱仪器中的电场与磁场。在质谱仪器科学里，有重要的Earnshaw定理：一块永久磁铁在恒定磁场中的势能只可能取到鞍点，而不可能具有最低点。换句话说，你永远不可能把一块永久磁铁稳定地放在另外一块永久磁铁之上。这个Earnshaw定理其实也是可以通过拉普拉斯方程来证明的。

(2)

拉普拉斯方程不但可以描述牛顿引力，还可以描述在一个弯曲面上的温度场，它也可以描述一个无旋流体的流体力学，也可以描述电子光

学的透镜的电场——如果有一个茶杯，它是铁的，均匀带电，现在问，这个茶杯周围的电势分布是什么样子的？一般来说，如果茶杯有一个旋转的轴对称性，那么我们可以推断，空间电势作为一个函数也是与旋转角无关的，电场的势函数分布满足拉普拉斯方程。但怎么计算这个茶杯附近的电场分布呢？必须要加上无限远处的边界条件才能求解。——尤其是当你用计算机做模拟计算的时候，你被迫先加上边界条件然后得到一个封闭区域，计算机才会做计算，问题是加什么样的边界条件。

因此，从拉普拉斯方程的角度来解读牛顿的万有引力，会发现如果要决定这个空间的引力场分布，必须要知道在无限远边界上的引力分布情况，但是，无限远在哪里呢？那里的引力势能分布一般被认为是零，这往往只对孤立体系才成立，对宇宙这样的系统，是不对的。这就是牛顿引力的另外一个疑难：边界疑难。

所以，在这个意义上我们可以了解什么是广义相对论？刻画广义相对论的爱因斯坦引力场方程并不是拉普拉斯方程，但它与拉普拉斯方程很像。只不过，因为在广义相对论中，很难找到一个类似于牛顿引力势的函数来做简单的刻画，所以我们需要修正这个拉普拉斯方程。那么，什么是牛顿引力势所满足的拉普拉斯方程，在广义相对论中应该如何对应呢？换句话说，在广义相对论中，拉普拉斯方程将被改写成什么样子呢？

在霍金和爱里斯的书《大尺度时空结构》中，他们也讨论了这个问题。霍金在1974年也就是他31岁的时候，写了这本天书。当他写这本书的时候，也许是为了想在人间留下痕迹，因为他患病了，医生说他快死了，所以，他把书写得非常之难。他本来就不指望这书能赚钱，没有想到，最后物理学家们还是很喜欢买下这本书放在书架上，以表明自己对引力也有追求。

为了节省篇幅，他们考虑一个简单的时空，就是这个时空中含有一个类时的凯林场（类时的意思是，这个矢量场的长度小于零，凯林场的

意思是说，这个是具有保度量的对称性的场，沿着这个矢量场走，你不会发现空间有任何变化）。这样的时空被称为稳态时空，就好像是一个已经热平衡了的水杯，这样的时空几何是比较稳定的。

在这样的时空，牛顿引力拉普拉斯方程可以被改写成微分几何的形式，并且据此可以大致看出爱因斯坦方程的来历。细节我们暂时就不写了，有兴趣的读者可以先看看《大尺度时空结构》，因为这个推导过程在那里就已经有了。

总之，可以把拉普拉斯方程写在弯曲的空间中，然后如果你足够懂微分几何，你也能慢慢顿悟出爱因斯坦场方程的形式来。不过这种事情一般人是不会的，因为一般的人，不知道怎么把拉普拉斯算子写成弯曲空间的版本——这要把度量写进微分里，有点儿复杂。

但是，在弯曲空间里，还有平均值定理吗？

你随便画一个球面，然后指望球面上的某个函数的平均值等于球心的函数数值，答案当然是没有这样好的结果了，但依然可以有一些类似的数学结论，这个就是数学家丘成桐曾经研究的问题了。

(3)

丘成桐1949年出生于广东汕头，老家在梅州蕉岭，在香港长大。他的父亲曾在香港中文大学的前身崇基学院任教。童年的丘成桐无忧无虑地生活。但在他14岁那年，他的父亲突然辞世，一家人顿时失去经济来源，所以丘成桐也就不得不一边打工一边学习。尽管如此，他仍以优异的成绩考入香港中文大学数学系。

他19岁的时候来到美国加州大学伯克利分校，20世纪70年代左右的伯克利分校是世界微分几何的中心，云集了许多优秀的几何学家和年轻学者。在这里，丘成桐得到IBM奖学金，并师从著名微分几何学家陈省身。

当时他的一位老师回忆他教过的这个学生说：“丘参加了我所教的一门课程，所以也可以说，我是他的老师。回想我第一次遇到他的情形，那时他冲进我的办公室，问了我一个关于负曲率流形基本群的问题。这个问题我从未考虑过，当然也就无从回答。不过我也感觉到，站在我面前的这个年轻人将会在数学上比我走得更远，这种直觉在1974年得到了验证，丘在Indiana Journal上发表了关于调和函数积分的文章（1976年刊出）。我称得上他的老师的时间恐怕只有一两年，自那以后我一生中成了他的学生。”

10年后他在世界数学家大会做1小时报告。当年他还只有28岁，也是在那一年，陈景润先生被邀请做45分钟的报告。这期间他证明了卡拉比猜想、正质量猜想，开创了一个崭新的领域：几何分析。

丘成桐取得博士学位后，在应邀前往普林斯顿高等研究院访问的一年中，他结识了许多年轻的世界一流数学家，完成了两篇论文。1972年秋，年仅23岁的丘成桐应邀来到纽约大学石溪分校担任副教授，又完成了几篇论文。在1973年美国数学会举行的微分几何大会上，丘成桐做了三个学术报告，以卓越的能力和杰出的贡献，向数学界显示了自己微分几何领域的领先水平。这一年是丘成桐数学事业上十分重要的一年，他完成了题为“完备黎曼流形上调和函数”的著名论文，用他自己的话说，这篇文章是他数学生涯的转折点。实际上，该文奠定了他应用分析方法的基本思想和技巧——所谓调和分析，就是解拉普拉斯方程。

外一篇 牛顿引力势能函数怎么从广义相对论那里得到

（较难，不具备数学基础的读者们可以跳过这一节）

牛顿在研究万有引力的时候，发现万有引力是和距离的平方成反比的，这就是著名的平方反比定律。换句话说，牛顿引力势能函数如下：

$$U(r) = -\frac{m}{r}$$

式中， m 表示的是恒星的质量。这个势能函数是满足三维空间上拉普拉斯方程的。

那么，如何把这个势能函数推广到爱因斯坦的广义相对论呢？

在史瓦西时空中，我们使用史瓦西坐标系，有如下度量：

$$ds^2 = -\left(1 - \frac{2m}{r}\right)dt^2 + \left(\frac{r}{r-2m}\right)dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)$$

这个时空存在一个类时的凯林矢量场，就是 $\left(\frac{\partial}{\partial t}\right)^a = K^a$ 。

此矢量场是时空的一个对称性（所谓对称性，在广义相对论中，指沿着这个矢量场生成的微分同胚变换是保度量的。）对称性是非常重要的，因为没有对称性就没有了很多有趣的物理量。但是，这个凯林矢量场 $\left(\frac{\partial}{\partial t}\right)^a = K^a$ 的长度并不归一，换句话说，它不具有单位长度。所以，它不是相对论中的所谓观察者，需要把它归一化，称为

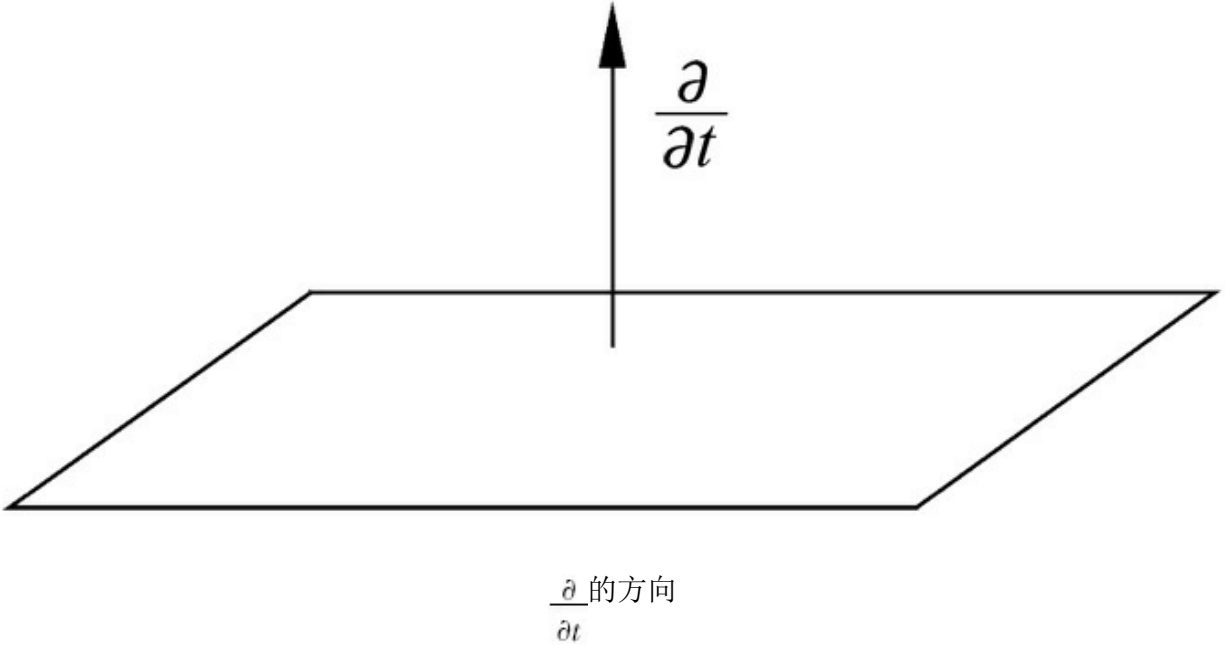
$$V^a = \left(\frac{\partial}{\partial t}\right)^a / \left|\left(\frac{\partial}{\partial t}\right)^a\right|$$

模长为

$$\left|\left(\frac{\partial}{\partial t}\right)^a\right| = \sqrt{1 - \frac{2m}{r}}$$

这样我们可以检验， $g_{ab} V^a V^b = -1$ 所以， V^a 可以是观察

者。



那么，我们引进一个深刻的问题，那就是，在史瓦西时空中，到底什么是一个静止的质点？这个问题很有意思，因为静止是相对于参考系来说的。只有在一个本身已经被我们研究清楚的参考系里，谈一个质点相对于这个参考系的运动才有意义。

V^a 是一个归一化的矢量场，它在空间每一点指定一个时间方向，是一个参考系。这个参考系是稳态的，意思等价于 V^a 乘上一个函数 f 以后，可以满足凯林方程。人们把这个函数 f 叫做红移因子。在不同的引力场半径处它有不同的值，很多广义相对论的效应都与这个函数 f 有关。

$$f = \left| \left(\frac{\partial}{\partial t} \right)^a \right| = \sqrt{1 - \frac{2m}{r}}$$

我们希望得到更好的性质，问一个问题， V^a 是不是超曲面正交的？

这个问题的计算并不复杂。我们先把 V^a 用度量拉下指标，做成一个度量对偶矢量，也被称为1形式场。

$$\omega_b = g_{ab} V^a$$

在这里，我们可以看出来， ω_b 只有时间分量，

$$\begin{aligned}\omega_b &= g_{ab} V^a = - \left(1 - \frac{2m}{r} \right) \left(\sqrt{1 - \frac{2m}{r}} \right)^{-1} (dt)_b \\ &= - \left(\sqrt{1 - \frac{2m}{r}} \right) (dt)_b\end{aligned}$$

我们的目标在于检验 V^a 的一个几何性质，就是问 V^a 是不是超曲面正交的，换句话说， V^a 在每一点是不是总与一个超曲面的法矢量平行。在这个要求中， V^a 不一定是这个超曲面的法矢量场，因为法矢量场的长度不一定是单位长度。（读者可以想像平面上点电荷产生的电场，其正交曲面是一个圆，这些圆构成一个等势能面，等势能面的法矢量大小并不是单位长度）这里面有一个Frobenius定理。这个定理只要检验， $\omega \wedge d\omega$ 是不是等于0。

先计算 $d\omega$ ，也就是上面提及的那个对偶矢量的外微分。

$$d\omega = - d \left(\sqrt{1 - \frac{2m}{r}} \right) \wedge (dt) \neq 0$$

进一步有

$$\omega \wedge d\omega = - \left(\sqrt{1 - \frac{2m}{r}} \right) (dt) \wedge d \left(- \sqrt{1 - \frac{2m}{r}} \right) \wedge (dt) = 0$$

以上等于0是因为出现 dt 的重复项。这是基于外微分的反对称性质。

所以，我们知道， V^a 是超曲面正交。这就说明， V^a 所决定的那个稳态参考系，还有更好的性质，它是静态参考系。超曲面正交的参考系是可以定义同时面的，也就是说能在史瓦西时空中很好地讨论爱因斯坦关于同时的观念。



史瓦西

因此， V^a 构成了一个静态的参考系，每一个观察者都离史瓦西黑洞的表面保持固定的空间距离不变（所以必须有外部的火箭推动这些观察者才可以使得他们保持静止，否则就会作自由落体运动，一头栽进黑洞表面，也就会掉出我们的史瓦西坐标系所覆盖的范围）。

从这个 V^a 之中，我们可以读出引力场的信息，最重要的信息之一，就是引力场的势能函数的精确表达式。

在这里，我们要先计算一下 V^a 的四维加速度 A^a 。这个加速度在一定程度上反映了 V^a 身上所带的那个使得自己保持静止的火箭的推动力。

根据定义， $A^b = V^a \Delta_a V^b$ 。

因为带有协变导数，强行计算会涉及克里斯多夫符号。但我们可以使用一定的技巧，来计算这个四维加速度。考虑到凯林场和归一化的性质，马上可以得到最后的结果是

$$A^a = \Delta^a (\ln f) = \frac{1}{f} \Delta^a f$$

我们发现 A^a 是一个全微分（这说明四加速场是可积的也就是 $dA=$

0，这个条件比超曲面正交要强一些），对我们有用的是

$$A_a = g_{ab} \Delta^b (\ln f) = d(\ln f)$$

我们称 $\ln f$ 是引力场的势能函数。它等于

$$\ln f = \ln \sqrt{1 - \frac{2m}{r}} = \frac{1}{2} \ln \left(1 - \frac{2m}{r} \right)$$

把 $\ln f$ 作泰勒展开，第一项就是牛顿引力势。

现在可以研究一下红移因子 f 的一个性质。

$$f = \sqrt{1 - \frac{2m}{r}}$$

先对四加速度的定义式两边故意乘上 f ，因为 fV^b 是凯林矢量场

$$fA^b = fV^a \Delta_a V^b = V^a \Delta_a (fV^b) - V^b V^a \Delta_a (f)$$

我们要利用四速度和四加速度正交的性质。就对上面式子与 V^b 缩并。

$$\begin{aligned} fA^b V_b &= V_b V^a \Delta_a (fV^b) - V_b V^b V^a \Delta_a (f) = 0 \\ V_b V^a \Delta_a (fV^b) + V^a \Delta_a (f) &= 0 + V^a \Delta_a (f) = 0 \end{aligned}$$

以上第一项为0是用到了凯林方程。

一个意外的小定理，稳态时空的红移因子 f 沿着 V^a 世界线是常数。

$$V^a \Delta_a (f) = 0$$

这一点在很多书上没有被强调，但在很多证明中需要用到，比如在霍金的学术专著《大尺度时空结构》中也是需要用到的。

以上计算表明 f 沿着观察者的世界线的方向是保持不变的一个常数。这个也符合直观，因为观察者距离黑洞中心的距离 r 保持不变，这也就是所谓静态的物理图像。但是，我们要问另外一个问题，那就是 f 随着空间方向的变化，它应该满足什么规律？因为我们已经知道， f 实

际上应该要满足一个类似于拉普拉斯方程的引力势能方程。

具体我们就不再详谈，有兴趣的读者可以参考《大尺度时空结构》一书。

24 三体问题

(1)

数学家庞加莱说：“人生是漫漫长夜里的灵光一闪，其他的毫无意义。”所以他觉得需要寻找意义，因为在1900年代他是在一线的当红数学家，所以考虑的问题自然是很重要的。他有很多彪炳青史的数学贡献，最不著名也是最重要的一个贡献，他发现了在椭圆曲线上的有理点（坐标都是有理数的点），这些点在曲线上的分布并不是连续的，而是一个一个的孤立点，而且这些点具有离散群结构。这个是近代数论的重要发现，而且数学家据此证明了350年来的数学谜语——费马大猜想。在这个基础上，人们想搞清楚一条椭圆曲线上的有理点的个数是不是有限的，如果是无限多个的话，那么这些点只有可以由几个基本的点来产生，这些基本点被称为椭圆曲线的秩。数学家把这个问题称为BSD猜想，如果谁可以解决就可以得到100万美元的奖金。闲话一句，本书作者自己也开发了一系列小软件，来研究这个BSD猜想。如下图所示。



让我们回到主题，牛顿时代以来，留下了丰富的遗产，那就是微分方程和万有引力。

庞加莱非常擅长研究微分方程，他渴望在最简单的相空间平面上找到封闭曲线来代表一个动力系统的演化轨迹，因为这对应了周期运动，也就是中国古代文化里的那种天道循环的宇宙观。相空间就是刻画物体运动的位置和速度的一个空间，在那个空间上指定一点，你就指定了运动所发生的位置和速度。在量子力学中，因为位置和速度是不可以同时确定的，所以，在量子力学的情形下，相空间中的每一点都是模糊的——这就好像在北京中关村的大街上可以看见很多大的LED电子显示屏，远远地看，显示屏上的每一个点都是清晰的，但走近了看则每一点都是模糊的。

庞加莱那个时代还不需要去处理量子力学问题，他那时候的数学问题，很多都是发生在经典相空间上的问题。他就好像一个孩子一样，在纸张上随便一画，他希望自己能画出一个封闭的平面曲线来，如果这平面是一个相空间的话。那么很明显，这个封闭曲线就对应的是周期运动（比如，女朋友走了以后，还会不会回来？这种问题在数学上是典型的判断周期性的问题。）这个宇宙需要和谐，周期运动是和谐宇宙的主旋律。潮起潮落，斗转星移，树木在夜风中婆娑，哪一个不是周期运动？可喜的是，庞加莱发现了，在（相）平面上，这样的周期运动的存在条件是非常简单的。这就是庞加莱-Bendixson定理，我们不再说细节，总之，如果是一个一维运动，那么它的相空间是二维的，所以，这个定理能帮助我们寻找到周期运动的解。

（2）

庞加莱当然也研究牛顿的万有引力，不过他不像玻耳兹曼那样研究

大量星体的统计行为，而是考虑3个星体之间的相互作用，这就是著名的三体问题。这相空间就不是二维的了，而是18维的，反正这是牛顿引力的最高境界了，这个问题总可以看成是3个星体的位置和动量（速度）在这个高维空间里的流动。庞加莱也是无利不起早，正好有一个比赛是研究这个问题可以得到奖金，于是庞加莱在这个问题上做了一些基础性的研究。

事情是这样的，在1887年，瑞典国王奥斯卡二世（1829—1907）悬赏2500克朗，征求天文学中一个重要问题的答案。这个问题就是“太阳系是稳定的吗？”其实这是牛顿本人早就提出来的一个老问题了。牛顿以当时已观测到的木星和土星运动的不规则性以及彗星以极扁的轨道横穿所有行星的公转轨道所可能带来的干扰作用为依据，提出了太阳系的运动可能会陷入紊乱的担心。此后不少科学家都对这个问题进行过探索。直到1784年，本书前面提到的那个数学家拉普拉斯根据万有引力理论证明，太阳系是一个完善的自行调节的机械机构，行星之间的相互影响和彗星等外来天体所造成的摄动，最终都会自行得到改正。所以，太阳系作为一个整体是稳定的，它将无限期地继续作目前的周期运动。但是看起来，拉普拉斯的答案并没有消除科学界的这个疑虑，没有阻止100年后瑞典国王的悬赏征文。

庞加莱也向奥斯卡国王的难题发起了进攻。但是，这个问题太困难了，它涉及了怎样研究一个复杂动力系统的稳定性这个深刻的问题——因为这表面上看是一个微分方程的问题，但实际上却与拓扑学也有关系，而当时的人都习惯于寻找方程的精确解，还没有那么图像化的思维方式——也就是定性分析的方法还完全没有萌芽。但是，这个新时代的杰出人物庞加莱没有像他的前辈们那样进行正面进攻，而是从侧面为了做这一研究创立了一个新的数学分支——拓扑学，拓扑学是研究一个气球如何被一个小孩揉捏的学问，只关心形状与会不会破裂这些问题，并不关心气球表面的弯曲程度。庞加莱的这个学问大大推进了人们对三

体这个历史难题的认识。在太阳系中，包含着十多个比月球大的巨大天体，这是造成解题困难的根本原因。如果太阳系仅仅由太阳和地球组成，这就是一个“二体系统”，问题则很简单，牛顿早已完全解决了它们的运动问题。它们的运动是简单而规则的周期运动，太阳和地球将围绕一个公共质心、以一年为周期永远运转下去；或者简化地说，地球将以太阳为一个焦点，周而复始地沿椭圆轨道绕转。然而，当增加一个相当大的天体后，这就成了一个“三体系统”，它们的运动问题就大大复杂化了，要彻底解决这个问题，几乎是不可能的。对短时间内的运动状态，可以用数值计算的方法来确定；但是由于根据牛顿力学所列出的方程组不能解析地求解——因为能找到的守恒量的个数小于方程的自由度，所以系统长时间的运动状态是无法确定的。

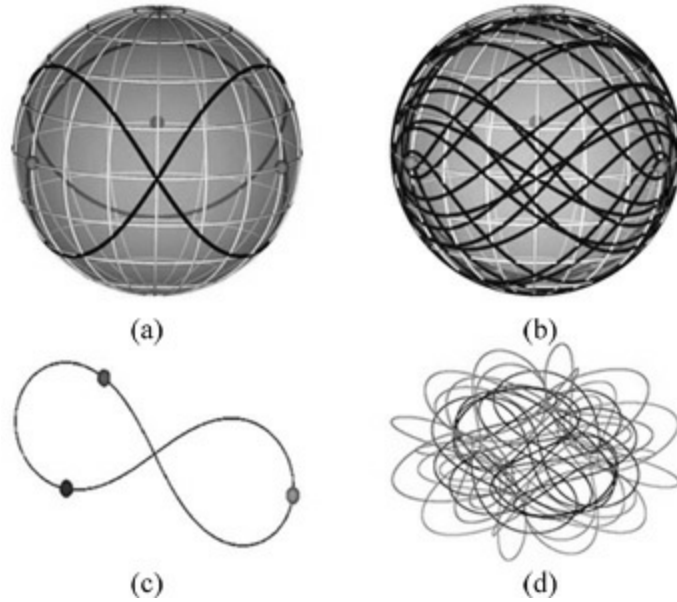
为了减少解决“三体问题”的难度，庞加莱着眼于美国数学家希尔（Hill, George William, 1838—1914）提出的一个极为简化的三体系统，即“希尔约化模型”。三体中有一个物体的质量非常小，它对其他两个天体不产生引力作用，就像由海王星、冥王星和一粒星际尘埃组成的一个宇宙体系一样。这两颗行星就像一个“二体系统”一样绕着它们的公共质心作周期运动；但这颗尘埃却受到两颗行星万有引力的作用，在两颗行星共同形成的旋转着的引力场中作复杂的轨道运动。这种运动不可能是周期的，也不可能是简单的，看上去简直是乱糟糟一团。为了用几何方法直观地描绘运动的情况，他用“相空间”来描绘运动过程。某一时刻系统的状态在相空间里用一个点表示；系统状态随时间的变化，即系统运动方程的解，对应于相空间的一条曲线，称为“相轨道”；如果物体作周期运动，它的相轨道就是一条闭合曲线；如果曲线不闭合，则表示物体的运动是非周期的。但是，为了确定系统的运动是不是周期性的，与其自始至终地跟踪系统运动的全过程，不如只观察系统的相轨道是否总会通过同一相点。设想通过相空间中一点A（初始状态）作一个横截面，如果系统的相轨道总在同一点A穿过截面，那么系统的运动就是周

期性用庞加莱截面考察运动情况的；相反，如果系统的相曲线1表示周期运动轨道每次都在不同点穿曲线2为非周期运动过这个截面，它的运动就是非周期的。这个截面现被称为“庞加莱截面”，它把对连续曲线（相轨道）的研究简化为对点的集合的研究，相当于对系统的全部运动过程进行不连续的抽样检验，从而简化了检测工作。

庞加莱还发现，在刚才那个空间里，可能存在一个不断回归到初始状态附近来的解，这个现象的发生被称为庞加莱重现定理。总之，他把要研究的问题写成动力系统的样子，然后找出这个映射下不变的测度，然后因为这个空间总的测度是有限的，所以随着时间的推移，系统的状态必须要回到初始状态的附近——这与玻耳兹曼的遍历假设有异曲同工之妙，但庞加莱重现定理只要求系统的状态能回到出发点的附近，而不是精确地回到出发点，换句话说，只要地球是有限大的，你运动的速度也不可以是无限大，那么你的相空间的测度（也就是体积）就是有限的，这个时候如果你从北京出发，周游世界，庞加莱重现定理会预言，你一定会回到北京附近，要么在唐山，要么在天津，要么在廊坊……不过庞加莱还是不能完全解答三体问题。

(3)

这确实是一个让物理学家日思夜想的问题，尤其是寻找三体系统中的周期性轨道运动。在最近很流行的科幻小说《三体》中，也将这个问题设置为整个故事情节所发生的时空背景，引起很多读者的兴趣。



三体问题的一些特解

自三体问题被发现以来的三百年中，人们只找到了三族周期性特解。最近，有两位物理学家完成了一个壮举：一口气找到了13族新的周期性特解。这项工作或许能帮助天文学家对行星系统有更深入的理解。要发现三体问题的周期性特解绝非易事。著名的数学家拉格朗日和欧拉在18世纪得到了一些结果——不过这些特解都是微扰不稳定的，风一吹就会崩溃。直到20世纪70年代，美国数学家Roger Broucke和德国天文学家Michel Hénon借助计算机又得到了更多的结果。他们把所有这些被发现的特解归结为下面3族：拉格朗日-欧拉族、Broucke-Hénon族和8字形族，拉格朗日-欧拉族的解非常简单，就是三个物体等间距地在圆轨道上运动，就像旋转木马那样。Broucke-Hénon族的解比较复杂，两个物体在里面横冲直撞，第三个物体在他们外围作环绕运动。而8字形族是物理学界Christopher Moore于1993年发现的——这就是上图（c）显示的那个解，非常精彩吧。8字形族之所以叫这个名字，是因为在这族特解中，三个物体在一条8字形的轨道上互相追逐。

在贝尔格莱德物理学院的物理学家Milovan Šuvakov和Veljko Dmitra Šinovic发现新的13族特解之后，三体问题特解的族数被扩充到了16族。

发现新的特解不是一件容易的事。三个物体在空间中的分布可以有无穷多种情况。必须找到合适的初始条件：位置、速度等，才能使系统在运动一段时间之后能够回到初始状态，即进行周期性的运动。Šuvakov和Veljko Dmitrašinović发表在《物理评论快报》上的论文描述了他们的方法：运用计算机模拟，先从一个已知的特解开始，然后不断地对其初始条件进行微小的调整，直到新的运动模式被发现。“我们所做的事情，思路非常简单，简单到大家都可以做”，Dmitrašinović说道，“当发现这些新的特解的时候，我们非常吃惊，然而更吃惊的是，这些解之前其他人居然没有发现”。

面对数量如此巨大的解，贝尔格莱德的两位物理学家发明了一种新的分类方法。他们运用了一种叫做“形状球”的抽象空间，通过物体之间的距离来描述轨道。拉格朗日-欧拉族解的最简单情形在这个球上就是一个点，因为在这个解中，物体间两两的距离是恒定的值。

然而其他的解就复杂得多了。比如被他们起名为“纱线”的解，在形状球空间中的形状就像一个线团，而在实际空间中，轨道的样子就像一大坨意大利面。Šuvakov和Veljko Dmitrašinović根据此方法把所有已发现的通解，包括前人发现的那些，总共分成了16族。他们又根据对称性和其他性质将这16族解分成了4大类，其中第一类囊括了所有前人发现的特解。他们下一步的工作，将会检测这些特解哪些是稳定的，也就是说即使受到小的微扰也能维持。这是关于三体问题最新的研究成果了。而爱因斯坦是在当时的三体问题刚开始火热起来的时代背景下成长的，他读大学是学物理的，他根本不想知道庞加莱等人所说的测度到底是什么。庞加莱等人是把牛顿引力升华到了一个相空间流形上来研究微分方程的流了，而爱因斯坦当时并不知道将来自己要把牛顿引力升华到另外一个境界。

25 惯性参考系

(1)

康托建立了集合论的基础，数学大厦的基础得到奠基。他和希尔伯特于是提出一个问题，他们猜测，无法构造出一个集合，使得这个集合的势在自然数集合的势与实数集合的势之间。这个问题被称为连续统假设，相当难。这个假设是对的吗？

过了好几个春秋，哥德尔和科恩才证明，这个问题无解。也就是说，集合论的公理体系与这个连续统假设之间是相互独立的，我们无法从集合论的公理体系中推出连续统假设到底是错的，还是对的。所以，以集合论为基础是数学的一个孤岛，而连续统猜想是另外一个孤岛。

在物理学上，也有这样的孤岛吗？在物理学上，也有一些问题非常困难，其中一个问题就是：牛顿关于惯性系的说法是对的吗？牛顿用他的第一定律定义了一个惯性系。牛顿提出，一个不受力的物体保持匀速直线运动，那么这个背景舞台就是一个惯性系。可是，牛顿怎么知道这个物体到底有没有受力呢？

牛顿时代以来，这个基本问题困扰着很多人，这好像是一个逻辑的死结，有的人因思考此问题而精神崩溃，有的人因此受到旁门左道的蛊惑，有的人则自己突破任督二脉而修得正果。

到底什么是惯性系？

惯性系就是这样一个孤岛。你不能证明牛顿关于惯性系的论述是对的还是错的，在牛顿理论的框架下，这个问题是无法解决的。我们暂时不要去追究牛顿的论述，不过物理学家可以构造出一个牛顿惯性系的、只能依赖于物理学实验的方法，在这些基础物理问题上，实验与测量把物理学与数学甚至哲学思辨区别开来。这就好像本书作者之一张轩中曾

经去北京水晶石教育给一些做动画的艺术家们讲过一次关于狭义相对论的科普讲座，在那次讲座上，张轩中给他们用世界线的方法讲了双胞胎悖论，这个双胞胎悖论最后归结为谁的四维加速度不等于零，张轩中最后不得不告诉那些艺术家们，这个问题只能通过实验来检验，谁更年轻那么谁的四维加速度就不等于零（关于这点可以参考本书后面的章节，此处不再赘述）。

(2)

关于惯性参考系的后来的事情还得从欧拉说起。欧拉虽然眼睛瞎了，但他心里像明镜似的，一尘不染。他活着的时候做了无数的事情，其中两件事情是与物理学有关的：

1. 写出了理想流体的运动方程。
2. 写出了绕质心旋转刚体的动力学方程。

欧拉关于刚体的运动方程，非常之重要。因为地球在自转的时候，地球没有受到外力矩的作用，所以，地球的运动就很好地满足欧拉的自由刚体运动方程。也许欧拉很清楚一件事情，那就是地球好像一个陀螺一样在太空中旋转，但它的极轴一直是永远指向北极星的——因为角动量守恒，并且因为极轴方向是地球作为一个刚体的转动惯量算子的一个本征方向（本征矢量是线性代数的基本内容，我们把转动惯量算子看成一个矩阵，那么它就有几个本征矢量所确定的方向）。很明显，一个自由的陀螺能够在太空中指定一个特定的方向，这个方向是不会变的。



用卫星精确定位地球上的位置需要建立一个惯性坐标系，并且考虑广义相对论效应

因此，利用3个自由陀螺，它们的自转轴方向指向远方的3个不共面的恒星，就能够建立起一个惯性坐标系来，这3个陀螺在牛顿力学的意义上，将指定固定的3个方向正好构成一个惯性参考系来。因此，可以说，欧拉的自由刚体使得牛顿的惯性参考系可以被构造出来。这在工程学上具有现实的意义。

自由刚体绕着转动惯量算子的3个本征方向旋转的时候，角速度方向和角动量方向是重合的。这3个方向分别沿着惯量椭球的3个坐标轴方向，所以，在构造牛顿惯性系的时候，我们需要让自由的陀螺绕着这3个本征方向的其中一个作自转，这样的话，这个陀螺绕着转的那个轴是随时间不变的。在2013年6月20日（星期四），中国的女航天员王亚平在“天宫一号”上给全国的中学生们研究了自由陀螺的刚体运动。所谓自由陀螺也就是不受到任何外力的陀螺。我们都知道，高速旋转的陀螺具有很好的定轴性，高速旋转陀螺的这一定轴特性在天上、地上是完全一样的。因此，有很多设备都是用陀螺组合来定向，那么，在“天宫一号”里，也安装了不少利用陀螺的定轴性原理制作的仪器，用来测量航天器的姿态。但是王亚平接下来也给大家演示在地面上是很难做到的，

不过在太空失重环境中就很容易实现的自由陀螺翻滚的现象。她把静止的陀螺放在那儿，给它一个干扰力，那个静止的陀螺在翻滚着向前运动，它的轴向发生了很大的改变——这个运动过程就可以用欧拉方程来描述。

在飞机、潜水艇中的导弹上，一般也可以安装这样旋转的陀螺，用它来指引一个特定的方向，这个方向在牛顿力学意义上是永远不变的。这就是惯性导航的大概意思。（当然，还有一种导航的方法是利用3颗卫星来作GPS导航，如果导航要做到很精确的程度，那么人们必须考虑广义相对论效应。）至少玩过google earth这个软件的人应该可以知道，在卫星上要精确定位你们家房子的位置，并不甚容易。至于另外一种导航的方法是使用所谓光纤陀螺，利用的是广义相对论中的Sagnac效应。Sagnac效应，最早的始作俑者，当然是Sagnac，此人是法国的实验物理学家，与居里夫人等人也算认识。他用牛顿力学的方式，解释了他发现的一个实验现象，那就是旋转圆周上背道而驰的光重逢以后会出现干涉条纹。

(3)

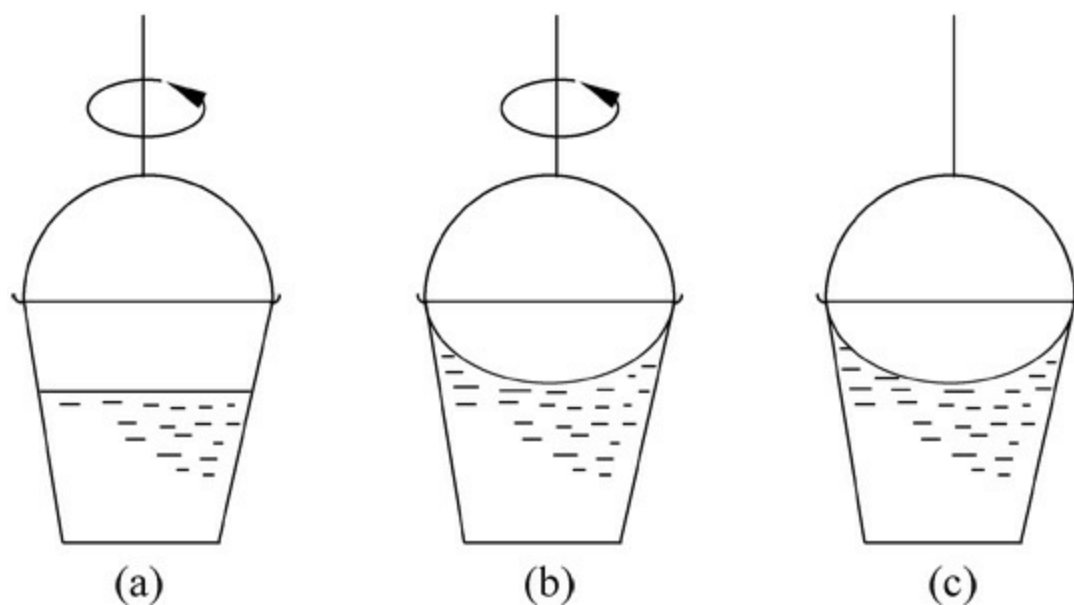
关于太空与惯性系统的隐秘联系一直牵动着世间千千万万人的心，但很多事情其实本质上是一样的。比如，有一个高僧，他曾经这样告诫自己的弟子：

不是风动，不是幡动，是你的心在动。

是的，虽然说万象皆是空妄的，但对于牛顿来说，到底什么在动，什么不动，是一个基本的问题。

1687年，牛顿45岁了，他的《自然哲学的数学原理》正式出版，在这本书中，牛顿用旋转水桶实验论证绝对空间的存在。他认为，旋转水桶中水面形状由平变凹，是由于水相对于绝对空间作加速运动而受到惯

性离心力作用的结果，水面变形正说明了绝对静止空间的存在。



旋转的水桶

牛顿的这种非常宏大的绝对静止空间，在相对论学家看来，是一个整体的惯性系。而实际上，这样整体的惯性系虽不是真实的物理，但也不失为一种理解问题的方式。在广义相对论中，可以有局部的惯性系，也就是说，我们可以在时空上的一个点上谈论惯性系。

牛顿看到水面的弯曲，认为这是相对于绝对空间运动而引起的。但是，在200年后，马赫认为，牛顿的说法是不对的，马赫说，水桶旋转的时候，水面变得不平坦，这是因为水在运动的过程中受到了全宇宙的其他物质对它的引力相互作用的拖拽。

马赫有这样一个相互作用的观念，他认为，水面的变化，一定是受到什么力的作用了。

马赫和牛顿的观念，其实在数学上来说，是等价的。正如在高中物理中，我们可以在非惯性系中加上一个虚拟的惯性力，使得运动发生在惯性系里。只不过，牛顿认为，惯性力是虚拟的，而马赫认为，惯性力是一种真实的力——就是引力。

这就是马赫的物理观念，在本书前面已经说过，马赫是一个很有物理追求的人，他第一强调的是观测，第二强调的是真实的力。所以，在这个意义上，马赫是反数学的，他没有数学倾向，他需要的物理，是真实存在的力和可以被观测到的现象。所以，马赫是一个很不错的物理学家。

在马赫的思想里，隐约可以看到，惯性力是和引力等价的。

仔细体会马赫的意思，可以发现，任何有质量的物体都与全宇宙的星体存在着相互作用。包括你在内，你每动一个小脚步，全宇宙的星体的位置都会变化。因为你们之间存在瞬间的非局部的引力相互作用。

而惯性就是这样产生的，别的星体显然是对你的运动有一种本能的抵抗，因为你动起来以后，它们也要被你带动，所以惯性是一种拖拽的引力效应。这是马赫的惯性观念。这个观念非常像电磁学中的楞次定律，那个定律描述线圈和磁铁之间的关系，当磁铁要插进线圈的时候，线圈阻碍磁铁插进去，当磁铁要拔出线圈的时候，线圈也阻碍磁铁拔出来。而惯性参考系也是一个与引力场纠结在一起的物理概念，它永远只能是一个近似，在物理上根本就没有那么理想化与那么纯粹的东西。

外一篇 从神十发射谈到电影《星际迷航》

“神舟十号”的升空把中国人也带进了一个航天时代，强化了普罗大众对太空的认知，而影片《星际迷航》系列可以说是普及相对论知识的绝好范本。

在“神舟十号”的发射后半程，宇航员手中的笔悬浮在空中。我们知道，在太空中的这种失重现象是最常发生的宏观物理现象，这主要是因为地球对卫星的万有引力全部用来提供卫星绕着地球旋转所需要的向心力，太空船就像地球的一颗小卫星，所以居于其内的人也是感受不到重力的。著名的物理学家伽利略最早对失重现象做了实验：他在意大利的

比萨斜塔上向地面扔下石头，发现石头在作自由落体运动，其运动时间与石头的重量无关。这是人类第一次在实验上去探索失重现象。

1907年，著名物理学家爱因斯坦在专利局做过专利审查员，他的狭义相对论已经发表，当他希望用狭义相对论来研究万有引力时，他重新思考了伽利略的自由落体运动，并且得到了一个惊人的发现：在自由下落的电梯里的人是感受不到重力的，因此，在自由下落的电梯里的人根本无法判断自己到底是不是在自由下落。根据这个洞察，深邃的爱因斯坦得到了涉及广义相对论的所谓“等效原理”。何谓“等效原理”，举个例子，将火箭中的航天员的眼睛蒙上，然后朝天上发射火箭，此时如果火箭转头向下从高空掉下来，这个时候作自由落体运动火箭中的航天员是感受不到重力的，他们也许会以为自己已经到了太空的无重力环境中去了。

这就是中国的航天事业应该给中国青少年带来的科学启迪，至少应该让现代的青少年知道，那个听上去高深莫测的爱因斯坦的广义相对论其实就是研究万有引力的。

再说到电影《星际迷航》，这个电影诞生于20世纪60年代的美国，当时美国与苏联争夺全球霸权，苏联在1957年第一次发射人造地球卫星以后，美国人感觉到自己的落后，不断努力，也开始发射卫星，然后美国人率先登月（当然也有人质疑美国人到底有没有真的登上月球）。在这一时代背景之下开拍的《星际迷航》系列在当时的观众中产生了强烈共鸣。此剧到2013年已拍摄12部。

著名相对论专家霍金也是一名《星际迷航》的爱好者，他甚至在电视剧版中还扮演过自己与牛顿和爱因斯坦一起打牌——虽然霍金的科学成就还难以达到爱因斯坦与牛顿的三分之一。

《星际迷航》属于硬科幻，里面的很多科学观点还是相当严谨的。

片中被广泛用于太空旅行的“曲速”（超光速飞行）是值得大书特书的。曲速，是在太空中作长距离旅行必须拥有的技术。因为人类所处的

太阳系，相对整个太空来说实在太渺小了。离太阳系最近的恒星系统是半人马座的“比邻星”，这个恒星离开太阳系的距离是4光年。刘慈欣的科幻小说《三体》中的三体人就居住在那里。从那里出发，光要走4年才能到达地球，而光的速度是一秒钟绕地球赤道走7圈。可见这星际距离实在很远，一般常规速度的航天飞机根本不可能以人寿到达那里。

那么，怎样才能快速到达瓦肯星这样距离地球16光年的行星呢？连光都要走16年，一般动力的航天飞机估计飞160年都飞不到，这时就需要“曲速”。

众所周知，根据爱因斯坦的狭义相对论，有质量的物体的运动速度不可能超过光速，这就好比你不加速你的汽车，到一定时候，你会发现汽车变得越来越重，最后无论如何也加速不了了，这是因为此时汽车的质量变成无限大了。再比如，以前电视机用的显像管，里面有电子枪通过高压打出电子来，随着电压的升高可以看到电子的速度在增加，同时电子的质量也在增加——这里就显示着狭义相对论的思想精髓：质量与速度相关。这个理论背后的前提是：我们讨论的是在一个“固定的空间背景下”的物体的运动速度。就好像空间本身是一个舞台，一个京剧演员站在台上演出，如果舞台本身固定，只有演员会运动，这个时候，无论演员怎么翻滚折腾，他的运动速度一定不会超过某个极限数值。

那么，“曲速”（超光速飞行）又是怎么回事呢？这个现象是由“空间本身的运动”引起的——根据爱因斯坦的广义相对论，空间本身也会随着时间变化，好比一个气球可以被吹大。关于空间随着时间会演化的观念是非常令人吃惊的，但是最近一百年来，天文学家已经观测到的宇宙空间正在膨胀——在我们银河系的左边与右边，其他星云都在离开银河系，这只能解释为宇宙空间本身在膨胀。而且膨胀速度是可以超过光速的——这就是哈勃定律。《星际迷航：暗黑无界》中的“曲速”也是利用了空间本身的膨胀：在时空中制造一个气泡来制造这种空间膨胀的效应，当然这在技术上实现起来是非常困难的。这时的空间是随着时间变

化的，就好像坐上了飞驰着的火车的蚂蚁，再怎么爬也没有火车本身开得快。这也好像剧院舞台突然倒塌（空间发生变化），那么演员本身随着舞台倒塌而下坠或翻滚的速度就可以超过演员本身的速度。

此剧中与相对论有关的情节还有一个，那就是史波克和未来自己（老史波克）的对话，神秘的时空穿越，有可能实现吗？现在不可能，将来，未必不可能。但是，在相对论科学中，这种穿越在逻辑上来说是不允许的——因为这样会改变历史。穿越回到过去会产生“弑母悖论”，假想一下，小王穿越回到20年前，碰见了母亲，当时的母亲只有13岁，还未结婚生子。小王失手将其杀死，那么这个小王是谁生的呢？

还有一个与相对论有关联的地方出现在第11部。第11部中提到的“红色物质”，这个红色物质很小的一块就能产生一个黑洞，将瓦肯星吞没。这个“红色物质”很有可能指的是中子团。中子团质量很大，就像中子星一样，大约1亿吨每立方厘米。我们知道，原子是由原子核与电子组成的，如果引力很强，把电子也压进了原子核，那么原子核中的质子就会与电子结合转化为中子。中子星拥有大质量和极小的体积，引力非常大，通过吸收周围的物质可以继续增大质量，直到一颗星球被吞完为止。中子星吸积周边物质的过程中有一个重要的概念叫做史瓦西半径，如果物质落入史瓦西半径之内，将再无逃出的可能，因为史瓦西半径之内就是黑洞的内部。这里的时间与空间对换——本来时间只能朝一个单一方向走，不存在回头路——在黑洞内部会变成空间只能朝一个方向走，没有回头路。这些是广义相对论的基本内容。

神十升空与《星际迷航》里面的科学知识非常庞杂，本文主要讲到的是涉及广义相对论的部分。

26 伯尔尼克拉姆大街49号

(1)

1903年1月6日，爱因斯坦和大学同学、出生于塞尔维亚的女物理学家米勒娃结婚。是年秋天，几经搬迁，他们最终选择租住在伯尔尼老城中心的克拉姆大街49号三层的公寓。

这个时候，爱因斯坦终于有了一个稳定的工作，有了一个家，如果仔细体味他经历的辛酸曲折，耳边会响起信乐团的歌《海阔天空》：

我曾怀疑我 走在沙漠中
从不结果 无论种什么梦
才张开翅膀 风却变沉默
习惯伤痛能不能 算收获
庆幸的是我 一直没回头
终于发现 真的是有绿洲
每把汗流了 生命变的厚重
走出沮丧才看见 新宇宙
海阔天空 在勇敢以后
要拿执着 将命运的锁打破
冷漠的人
谢谢你们曾经看轻我
让我不低头更精彩的活

凌晨的窗口 失眠整夜以后
看着黎明 从云里抬起了头
日落是沉潜 日出是成熟

只要是光 一定会灿烂的
海阔天空 狂风暴雨以后
转过头 对旧心酸一笑而过
最懂我的人
谢谢一路默默地陪我
让我拥有好故事可以说
看未来 一步步来了

要知道，在爱因斯坦大学毕业以后的两年里，他没有找到一个正经的工作，还曾经有一段依靠做家教来维持生计。1902年2月5日出现在瑞士《伯尔尼都市报》的一则广告如下：

1902年2月5日《伯尔尼城市报》启示：

...den bündischen Nr. 4978 ...he 7a, 2. St.	Ges. von Wälchen für Hiniographen.	Lad
Mädchen, ...ochen kann, sucht ...5049 ...empfiehlt sich als	Vermischtes Privatkunden in Mathematik u. Physik für Studierende und Schüler erteilt gründlichst Albert Einstein Inhaber des eidgen. ...Schleibersdiploms.	mit zwei Schauten ...vier Jahren Geschäft an g mit gutem Erfolg ist wegen Ueberne Geschäftes zu ü Gefl. schriftlich Nr. 5028 an da
derin. ...im g ...ung ...laue ...al	阿尔伯特·爱因斯坦，愿私人为大学生或者中学生 彻底 讲授数学和物理学。本人持有苏黎世联邦工学院的教师资格证书，住正义街32号一楼。试听免费。(FLI)	

所以，现在的工作与生活的稳定看起来还是来之不易的。爱因斯坦白天在专利局上班，晚上在家里做物理，妻子很快又生了一个孩子。因为专利局的工作甚是清闲，使得他有大量时间思考物理学。

“老婆，我真的很感谢格罗斯曼同学，他介绍的这个工作真是不错。”爱因斯坦说。

“对，我们应该有感恩的心。格罗斯曼是我们生命中的贵人。你现在有这样的成就，我真的好高兴。”米勒娃说，对她来说，能在一个陌

生的城市找到属于自己的男人和一个家，是很好的结局。

爱因斯坦有超强的物理直觉，他时刻受到一种天人感应式的启示，在用马尔可夫过程开始做出布朗运动后，他还开始思考光电效应——这个研究成果被广泛应用于数码照相机与电动自动旋转门等各个领域。最重要的工作也马上诞生，那就是狭义相对论。

狭义相对论和量子力学是两门不搭界的学问，也就是说，如果相对论是错的，量子力学也可能是对的，因为量子力学处理微观的物体，而狭义相对论则是处理高速运动的物体。那么到底什么是狭义相对论呢？本书的作者之一张轩中曾经写过一本关于相对论的科普书《相对论通俗演义》，在书中张轩中写了一段古文简单地介绍了相对论思想的发展：

人类历史，始于蒙昧。置身悬崖，若无必死之心，望而却步。引力做功，碎骨粉身。虽引力乃最弱之力量，然其统治大宇宙之脉搏呼吸。亚里士多德指出引力秘密，为世人所奉。而千年之下，伽利略用思想实验，得自由落体真谛。此情此景，历历在目。几百年来，比萨斜塔，默默无语。上帝造物，宇宙洪荒，乃至牛顿，方显露一丝天机。牛顿引力，平方反比。方程曼妙，让人叹为观止。

牛氏遗腹，少年孤冷，其引力方程，上穷碧落，下至黄泉。玉兔升空，苹果落地，牛顿方程，半窥天机。方程横空出世，世间犹惊。返谷溯源，乃开普勒之鞠躬尽瘁。白日西匿，开普勒于月牙高台，夜观天象，夙夜不眠。经济窘迫而身形憔悴，然于故纸之间，发行星运动之三定理。此乃呕心沥血之作也。

红尘百草，阡陌凋零，牛顿之后百年，法拉第铁匠之后，瓮塞绳黍之子，其英雄气长，研习安培之环路电流。奥斯特发现电可成磁，法拉第十年一剑，磁能生电。此间十年，实验无数，遍尝失败。

电磁相生，法拉第立场论。隔四十年，麦克斯韦设位移电流，写积分方程，电磁统一大成。然方程之中，未现磁单极子。狄拉克沉默寡言，深思熟虑，考虑量子力学，得电磁对偶。若单极存在，焉能以单一磁势覆盖球面？此乃纤维丛也。

场论既成，天下太平。爱因斯坦，犹太之子，叛逆之徒，其遍习引力历史，深通场论之奥义。厚积薄发，终成绝响。麦氏方程，天生超越伽利略之单纯空间变换。闵可夫斯基，乃得四

维时空变换，保持麦氏方程。庞加莱亦有所闻，一时间群雄并起，狭义相对论，若隐若现，爱因斯坦，确凿物理实在，于线性时空变换下不变。若较爱因斯坦之于庞加莱，庞氏深谙数学，而爱氏直指背后物理，深邃迥异于常人。此西元1905，爱因斯坦，声名鹊起于天下，此洞见时间须于空间一起，共铸时空流形，此乃千年来最广博深远之发见，天下唯此一人。

历史机缘巧合，全在冥冥之中。狭义相对论已大白于天下，然世间芸芸众生，依然不知所云。所谓动尺收缩，此中可见牛顿引力方程，水火不容于狭义相对论。牛顿引力，描述空间两点之距离，然空间两点之距离不复绝对，牛顿引力，势必依赖于观察之士。物理勿须依赖于观察之士，乃广义相对论性原理，此为物理之基本也。爱因斯坦，深味其中悲哀。方十一年，其演习黎曼几何，工于张量计算。友士格罗斯曼，居功至伟。数学家推动物理学之发展，此一例也。爱因斯坦日复一日，确信物理之规律，为人力所不能撼动，须臾之间，乃相信物理之规律，应于人人平等。其思辨于升降机中遇见等效原理，偶遇一生之最快慰思想。自由落体，于其深邃双眸，复现璀璨光明。

物理于几何，一衣带水。黎曼几何，天才之手，问世三十余年，遂转入爱因斯坦，发扬光大，天下人膜拜而熟习之。爱因斯坦引力方程，几何等于物理，方程美妙绝伦，意味隽永。此西元1915，希尔伯特，殊途同归，翁乃数学界之泰山北斗。引力历史于数学物理两家携手推动，此后数十年，群雄并起，爱因斯坦方程未绝于数学物理，虽人迹罕至，丛生皆云高山仰止，引力冷艳孤傲。

人生不相见，动若参商，然爱翁之于鸿蒙洪荒之回响，余音绕梁。

之后数十年，黑洞兴起，宇宙加速膨胀，一切归于漫漫黑暗。

早在16岁时，爱因斯坦就了解到光是电磁波，他想，如果一个人以光速运动，他看到的世界会是一个什么样子？比如在远处有一个钟楼，钟楼上的大笨钟的指针会走动，从远处看，这是因为钟上有光射到人的眼睛里来。现在假设小爱因斯坦以光速离开，那么从大笨钟发出来的光就追不上小爱因斯坦了，因此在小爱因斯坦看来，钟的指针就是静止不动了，时间好像就停止了。

爱因斯坦的少年时代一直困惑于这个问题，这个很有画面感的物理图像一直引导着他前进，后来使得他博得了冷酷历史的嫣然一笑。

这个问题把光的速度与时间的本质结合起来加以考虑，则是非常有趣的。

狭义相对论最基本的假设在于光速是一切速度的极限。当然这在相对论中显得非常自然，但却出乎普通大众的意料。换句话说，假如你认定在任何惯性系里，光总是以不变的光速运动，那么你也可以写出这些惯性系之间存在某种神秘的联系，这个联系显然不是伽利略变换，而是洛伦兹变换。

狭义相对论的最主要的公式是洛伦兹变换，是洛伦兹最先给出的，但相对论的创始人却不是洛伦兹而是爱因斯坦。洛伦兹也认为，相对论是爱因斯坦提出的。

从麦克斯韦那里可以知道，电磁波以光速传播，而且光速是一个恒定的常数。伽利略相对性原理说，物理规律在一切惯性系中都是相同的。麦克斯韦方程组在所有惯性系中都应成立，这就是说，光速在任何惯性系中都应该相同，都应是同一个常数 c 。但同时按照伽利略相对性，惯性系之间可以差一个相对运动速度 v 。依照速度（矢量）叠加的平行四边形法则，电磁波（即光波）的速度如果在惯性系A中是 c ，那么，在相对于A以速度 v 运动的另一个惯性系B中，就不应再是 c 了，而应是 $c+v$ 或 $c-v$ 。但是，麦克斯韦电磁理论说光速只能是 c ，不能是 $c+v$ 或 $c-v$ 。于是，爱因斯坦意识到，一定有什么地方出错了。

下面的三条理论，肯定有某一条是错误的了。

（1）麦克斯韦电磁理论，它要求光速只能是常数 c 。

（2）相对性原理，它要求包括电磁理论在内的所有物理规律在一切惯性系中都相同。

（3）伽利略变换，作为三维空间矢量叠加原理的平行四边形法则。

爱因斯坦觉得，第（3）点不能推广到四维时空。对于一个四维矢量，这个平行四边形法则不能继续使用。对于四维矢量的合成，必然会

导致一个新的数学发现，那就是数学家哈密顿曾经研究过的四元数。爱因斯坦当时不可能从四元数的角度来思考问题，正如现代的物理学家不太懂得代数几何一样，用不着，就没有必要去懂它，物理学家中有像狄拉克那样的，他可以自己发明一些高超而出人意料的类似于神来之笔的数学。但当时的爱因斯坦认定，第（3）条是错的，在光速不变原理和相对性原理的基础上，他推出了两个惯性系之间的坐标变换关系，这个关系就是洛伦兹等人早已得出的变换公式。

不过，爱因斯坦是在不知道洛伦兹等人的工作的情况下，独立推出这一公式的。更重要的是，爱因斯坦对该变换的解释与洛伦兹完全不同。在物理解释上，洛伦兹认为这个变换是运动参考系和绝对静止参考系之间的变换，而爱因斯坦是正确的——爱因斯坦认为，这是任意两个惯性参考系之间的变换，根本就不存在什么绝对静止参考系。于是，狭义相对性原理出场了：“所有的惯性参考系中，物理规律是一样的。”

这时候，时间停留在1905年，爱因斯坦大学毕业的第5年。

在这一年，爱因斯坦26岁，额头上并未长有菱角。但他表现的天赋异秉，确凿似乎已经超越了他的时代。

因此，简单地说，爱因斯坦的狭义相对论是这样说的：

1. 有静止质量的物体，它的运动速度在不同参考系中看起来是会变化的。
2. 无静止质量的物体，它的运动速度在不同参考系中看起来是不会变化的，都是光速。

（2）

狭义相对论把光速与时间的关系进行了澄清，也就是说，以光速运动的物体它本身所带的时钟是停止的。所以光本身是不能来刻画时间的，换句话说，光对时间来说就好像一把没有刻度的直尺。那么时间到

底是什么呢？其实，任何有质量的物体本身都是一把刻画时间的卷尺。这就是所谓“世界线”的概念，世界线的长度就是一个物体本身的真实时间——在物理学上叫做proper time，翻译为“固有时间”——就好像我们说钓鱼岛是中国的“固有”领土一样。要谈论“世界线”，首先必须有一个正确的时空观念。那就是要掌握狭义相对论最核心的思想：把时间和空间结合在一起作为一个整体来考察。比如，一个孕妇在怀孕的时候，可以选择做二维B超，也可以选择做三维B超去观察肚子里胎儿的情况：二维是平面图像，三维则是立体图像。而最新近的四维B超，则就是三维立体图像加上一维时间组成的动态录像，可以观察到孩子在孕妇肚子中吸吮手指、抓耳挠腮等可爱的动作。这动态的录像，就是四维的，把时间也加进来了，与空间信息组成一个整体的考察对象。爱因斯坦当然也是在物理学中加上了时间因素，而且把时间不再作为只具有唯一前进方向的客体：不同的人，有不同的时间——这个时间箭头的方向，由人运动的速度与光速一起决定。时间是私人的，就好像我们每个人都有电脑，都有手表和手机，但这些时间走动的快慢往往是不一样的（因为不同的电子产品里面表征时间走速的晶振不一样），至少我看到的时间和你所看到的时间不可能是精确同步的。

在这个意义上，火车的列车时刻表是没有意义的，因为根本不存在公共的标准的时间，不同的旅客都有不同的时间走动的速率。

所以与财产是私有的一样，时间的产权性质也是私有的。公共时间在物理学上是一个错误的概念，就好像公有制不能促进生产力大发展一样。

爱因斯坦一个人在思考。可惜，爱因斯坦感觉到自己的几何学知识，完全不够。他不知道如何来描述一根被引力场弯曲的世界线。

如果从更加宏大的视野来看物理，爱因斯坦所遇到的问题，需要一门叫微分几何的数学学问，需要一个数学概念，那就是“弯曲流形”。世界线可以看成是时空流形上的曲线。

对于当时的爱因斯坦来说，流形还是一个前沿的数学概念，整个物理学界对数学是排斥的。物理学家不会在乎数学家到底在做什么，除非有人能把数学半岛和物理半岛之间的桥梁建立起来。

当时物理学家的基本数学水平，就是经典传统的矢量分析。矢量分析是平坦空间上的一些矢量场的微分和积分运算。如果读者们依然有宏大的眼光，那么简单地说，这一套东西其实可以概括为一个英国诺丁汉面包师的工作。这个面包师傅，就是格林。格林小学还没有毕业，就去帮忙和父亲一起做面包了。这有点像当年江苏的一个小杂货店里的华罗庚。

一个人如果没有理想，那么和咸鱼没有区别。

当时的面包师傅格林也一样，他看到面包已经反胃，他心想：“难道我一辈子就这样了吗？我要做矢量分析。”过了很久，等他父亲死了，他卖掉了面包店，开始去大学读书。他心里早已经有了很重要的数学结论。作为一个有谱青年，他的格林公式说：“矢量场沿着一个封闭曲线积分，等于这个矢量场的微分在封闭曲线的内部积分。”

他的结论可以被推广到高维。他自己本人是做了二维和三维。当然，无论格林有多么牛，他无法超越时代，如果换成现代的数学语言来说，格林公式就是微分几何里最基础的对偶定理。

伯尔尼克拉姆大街49号的爱因斯坦斜倚在床上，他在纸上画啊画。他想把世界线也写成矢量的积分曲线，然后让世界线弯曲起来。可惜他的数学水平，和一百年前的面包师傅是一样的。爱因斯坦很是苦恼：时间是私人的，世界线是弯曲的，到底要怎么来描述引力对世界线的影响

呢？世界线的长度是每一个私人的固有的时间观念，这个才是本质的，但因为引力把世界线弯曲以后，使得每一个个体的私人时间更加不可能被同步校准到一个公共时间，所以，在相对论的世界里，不但不存在精确的全国统一的北京时间，而且引力场似乎要影响到每个人的时空观念？那么引力是如何与时空发生耦合的呢？

27 公务员的奋斗

(1)

1907年，爱因斯坦还是伯尔尼专利局的职员，他是一个公务员。他有一个叫贝索（Michele Besso）的哥儿们，也在这里上班。他经常和贝索一起讨论学术问题。他们两个人算是公务员系统中出现的少数几个真正有物理思想的青年，犹记得1905年那篇名为“On the Eelectrodynamics of Moving Bodies”的文章在《德国物理学纪事》（Annalen der Physik）发表的时候，此文章没有任何引用文献，只是在结尾的时候感谢了一个人。写这篇文章的人叫爱因斯坦，被感谢的人就是贝索。

爱因斯坦凭借在固体比热、布朗运动、光电效应、狭义相对论、广义相对论、激光原理、玻色凝聚、EPR佯谬、量子力学基础的完善、哲学问题的思考以及对诸多后辈物理学家的指导和大量的学术和政治活动获得了无与伦比的名声和荣誉——虽然他自己并不认为这是恰当的，而贝索除了在爱因斯坦的文章里出现过一次，再也没有什么能被别人记住的东西——他也许仅仅是爱因斯坦生命中的一个客星，甚至似乎没有真正地闪亮过，也没有真正地被人们记住过。他没有出现在其他名人的传记里，没有量子力学史或者物理学史记载过他，没有人为他作传，在爱因斯坦成为学术明星的时候，他还在伯尔尼专利局从事那些肯定让他有些无聊的工作。

其实，贝索与爱因斯坦在1896年相遇于一次音乐会上，他长爱因斯坦5岁。贝索是个多才多艺、知识渊博、兴趣广泛而且坚持学习的人，曾研究过物理学、商法、民法、生理学、英国文学、天体力学，后来在1908—1909年冬，爱因斯坦在伯尔尼大学（与专利局在同一个城市）任编外讲师（没有薪水的！）讲放射理论课的时候，全班只有两个学生，

其中之一便是贝索。正是他向爱因斯坦介绍了马赫的《力学史》，而马赫正是被爱因斯坦称为相对论先驱的人，虽然马赫本人强烈否定自己与相对论有任何的关系。而后，爱因斯坦将其房东家的女儿安娜介绍给贝索，后二人结为夫妻——同时值得注意的是，爱因斯坦与安娜的妹妹谈过恋爱。

贝索的活动并没有太多记载，通过梳理爱因斯坦给他的信件，也许可以看到贝索与爱因斯坦的关系和他在爱因斯坦心中的地位以及他一生的主要工作。

贝索保留的爱因斯坦的第一封信是1903年爱因斯坦在伯尔尼专利局工作时写给贝索的，当时爱因斯坦的妹妹Maja在贝索家里做贝索妹妹的家庭教师：

.....我的论文结果多次重写和修改之后，终于在星期一寄出了.....上星期Miza得了感冒，现在我也传染上了。今天无法去专利局了，不过已经好转，因此明天我就要上班了。

.....先父的老会计居然能做到使我善良的妹妹Maja相信他在道义上有义务把她的大部分收入交给他.....我已经在这个人几次撒谎的当场抓住了他.....他在我身上也曾试过.....愿上帝宽恕我，我以为我多少还是一个清醒的人，不易任人摆布.....

.....现在我重新决定去担任编外讲师，倘若能够实现的话。此外，我将不考博士学位，因为这对我没有什么帮助.....我最近将从事研究气体中的分子力，然后全面地研究电子理论.....

这封信的内容包括自己的家事、生活和自己的思考与打算等，从这些只言片语可以看出他们的友谊之深厚。

1904年爱因斯坦推荐贝索为专利局咨询工程师，从此两位朝夕相处达五年之久，爱因斯坦曾经回忆：

.....专利局把我们结合在一起，我们下班途中的谈话引人入胜，无与伦比，人事浮沉对于我们似乎不存在。

.....在整个欧洲，我找不出一个比他更好的知音。

.....他的成就只能在他所造就的人中找到。

等爱因斯坦后来成为大物理学家以后，爱因斯坦也从来没有将贝索

当成一个不懂物理的民科，每每有新的思考，都会与贝索通信讨论，下面是部分信件的节选：

〔量子力学〕近来工作没有多大成果，最有兴趣的是我发现了符合麦克斯韦方程的能量分布有无穷多种，也许量子问题的答案最后就隐藏在这里..... 〔1909年12月31日

于苏黎士〕

〔固体热传导〕目前，我正在试图用量子假说推导出固体电解质中的热传导定律
〔1911年5月13日于布拉格〕

〔第一届索尔维会议〕在电子理论方面，我没有进展，在布鲁塞尔，人们怀着悲伤的情绪看到这个理论的失败，找不到补救方法，那里的大会简直像耶路撒冷废墟上的悲号 〔1911年5月13日于布拉格〕

〔引力理论〕我已经严格证明，一个闭合的静止体系总能量——包括引力场能量——既决定它的惯性质量，也决定它的引力质量，目前我正为量子论问题大伤脑筋，但是成功的希望不大..... 〔1913年年底于苏黎士〕

〔引力理论〕我通过一个简单的计算已经证明，这些引力方程对于一切满足这个条件的参考系都能成立.....惯性质量和引力质量以及引力场的严格等效性，我记得在你来访时，我已经证明过.....现在我非常满意，不管对日蚀的观测成功与否，我对于整个体系的正确性已经不再怀疑.....我将住在大莱姆.....希望你尽快来看我..... 〔1914年3月于苏黎士〕

〔引力理论〕今天我已经把论文寄给你，如今实现了最大胆的梦想：普遍协变性.....
〔1915年12月10日于柏林威玛村〕

〔引力理论〕现在，连普朗克也开始重视这件事了.....这一切我以后跟你谈，但愿我们很快能相见 〔1915年12月21日于柏林威玛村〕

〔引力理论〕引力方面获得巨大成功使我非常高兴.....研究闵可夫斯基对你不会有什么帮助，他的论著是无用的复杂..... 〔1916年1月3日〕

〔激光原理〕关于辐射的发展和吸收，我突然有所领悟.....这是普朗克公式的直接结果，这一切全是量子的，我正在把结果写成文章..... 〔1916年8月11日〕

〔引力理论〕在那里，广义相对论已经获得了极大的活力.....我和埃伦菲斯特，尤其是和洛伦兹度过了一些难忘的时刻.....我终于放弃了离婚的念头.....空间和实践的客观意义首

先在于..... [1916年10月31日]

[统一场论]我已发现一种几何学，它不仅有黎曼度规，而且还有绝对平行性.....我一定很快把这篇论文寄给你，如果你不伸舌头，你就是个伪善之徒..... [1929年1月5日]

[统一场论]微妙之处在于：在四维空间中引入了五维矢量..... [1931年10月30日]

[统一场论]你一定知道，迄今的尝试全都已经失败.....我一旦有了哪怕是微小的有根据的信念，我就很愿意告诉你。 [1942年8月]

[量子力学哲学基础]我对统计性量子理论的反感不是针对他的定量的内容.....你是否注意到泡利对此的回答多么不合逻辑？ [1949年7月24日]

[统一场论]以统计学为基础的理论，尽管取得了很大的成功，但还是停留在事物的表面.....不幸的是，从逻辑上看来很简单的事情计算起来却很复杂..... [1949年8月16日]

[量子力学]五十年的思考还不能回答光量子是什么..... [1951年12月12日]

[EPR佯谬]由一个函数所描述的态和一个一定实在情况有什么关系呢？这样实在的状态就不是可以有经验直接感知的.....我拒绝这种假定.....证明如下..... [1952年10月8日]

[引力理论]你明显把自己放在一个不稳定的地方.....解释时间箭头的全部问题同相对论问题毫不相干.....爱丁顿的论证也许可能有些真实的东西，但是他缺乏批判精神..... [1953年7月29日]

[引力理论]广义相对论的话是完全正确的.....但是到现在为止还不能证实.....我们大概不能亲眼看到他的结局了..... [1953年9月22日]

[引力理论] 你对广义相对论的阐述很好地说明了它的发展史方面.....对牛顿基本原理提出的这个反对意见的核心..... [1954年8月10日]

1926年，贝索因为业务过少险些被伯尔尼专利局开除，爱因斯坦写信给专利局，对贝索有所评价：

“.....他有敏锐的理解力，他不喜欢斗争，因此他对每一个可供选择的方案都提出一个问题.....

他随时准备了解别人的见解，从中找出问题的焦点，因此他在讨论中很出色.....他能够成为卓有成效的批评家.....像苏格拉底一样.....

.....直截了当地说，他是一个卓越的非专家.....他是一个典型的教导者，而不是一个培育者.....

凡是对于他遇到的每一个向他求教的人有用的东西，对于他自己却是有害的，因为它永远不满足于已有的东西，没有什么论文署他的名，他的成就只有在他所造就的人当中才能找到。”

没有什么人评价过贝索，看上去他是不幸的，但是爱因斯坦给了他如此高的评价，这又是贝索的幸运，贝索的不幸在于他没能将自己的思考转化成结果，贝索的幸运在于他的思考居然能帮助一个几乎是历史上最伟大的人物。

贝索于1955年3月15日死于脑血栓，而爱因斯坦死于1955年4月18日，这这也是一个巧合，伟大的友谊终于随着生命结束而结束了。鲁迅先生曾经说过：在未有天才之前，先做培养天才的土壤。在我们为爱因斯坦的成功而顶礼膜拜的时候，是否应该了解那土壤呢？爱因斯坦当时在专利局作一个无聊的公务员奋斗的时候，是谁陪在他身边与他一起前进呢？

(2)

1907年有人请爱因斯坦写一个介绍狭义相对论的综述文章，写这样的文章，使得爱因斯坦重新全面地审视了一下自己的理论和周围的世界，他发现私人的时间观念不可以同步到一个公共的标准时间，这就好像钟表店的老板永远不可以把他店的钟表全部校准一样，这背后的原因在于，根本不存在一个全局的惯性参考系.....



任何一家钟表店的钟表都不可能精确对准，在相对论中也一样

而在现实中，一样是惨淡经营，爱因斯坦关于狭义相对论的文章发表以后，江湖上反应非常冷淡。这种冷淡让爱因斯坦有点沮丧，他知道自己在专利局而不是在大学的身份使得自己看上去非常像一个民间科学家。

其实，在柏林大学，普朗克的讨论班里，就在开始讨论爱因斯坦那个离经叛道的狭义相对论——只不过是爱因斯坦本人不知道而已，讨论班有一个人没听懂到底什么是狭义相对论，但他相信这一定是一个重要的学问。这个人就是劳厄，他打定主意，一放假就去伯尔尼大学向“爱因斯坦教授”求教。

风餐露宿，赶路匆忙，劳厄赶到了伯尔尼大学，问道：“我想拜访你们这里的爱因斯坦教授。”

伯尔尼大学的人说：“爱因斯坦？教授？谁？”

幸亏劳厄不是一个愣头儿青，马上说：“啊，他不是伯尔尼大学的，那他……我只知道他在伯尔尼。”

伯尔尼大学的人说：“这个人是要干什么的？”

劳厄说：“他很神奇的……好像是公务员吧。”

伯尔尼大学的人说：“他是公务员？你去专利局看看吧，那里好像有一个人叫爱因斯坦。”

劳厄说：“谢谢。”

赶到专利局，劳厄一头撞上一个年轻人，穿得像一个“油”博士，在走廊里来回踱步做思考状：一件不太干净的格子衬衫，领子已经不再坚硬，耷拉的脑袋上头发乱成一团麻，眼神甚是迷离，似乎有点忧郁。

专利局的走廊上空荡荡的。

“爱因斯坦博士在哪个办公室？”劳厄问。

年轻人愣住了，不知道说什么好，想了半天，说：“我是爱因斯坦，对不起，我们这里好像没有第二个爱因斯坦。”

劳厄惊讶了，但很快就笑了，说：“你好，我是来请教你问题的。”

爱因斯坦说：“你好，请教什么？”

劳厄说：“外面有一家小饭馆，如果可以，我们是不是边吃边聊？”

爱因斯坦说：“可以。”

几分钟后，两人在一家小餐馆的桌旁坐了下来。

“我是普朗克教授的学生。我们对你的狭义相对论很有兴趣。”劳厄说，“我也觉得，我们的时空观需要被革新。”

“你是做什么方向的研究的？”爱因斯坦说。

“光的干涉什么的，不过我最近对X射线衍射也有兴趣。”劳厄说。

.....

爱因斯坦与他相谈甚欢。

劳厄走了，过了几年，他做了X射线对晶体的衍射实验，使得人们进入到了一个真正实用的物理领域：X射线衍射分析。但劳厄对狭义相对论还是如痴如醉的，写了地球上第一本相对论方向的书《相对论原理》。爱因斯坦的生活没有被劳厄所改变，不过这次学院派的人的来访，增加了爱因斯坦的信心，他觉得自己还是有药可救的。当下最要紧的事情是去大学或者中学当个老师，这样才可以最后成为教授。

(3)

专利局的日子，什么时候才能到头？1907年，爱因斯坦思考的是如何把世界线和万有引力结合起来。但一口吃不成胖子，他觉得先思考一下万有引力。这一次思考爱因斯坦发现了前文提到的等效原理：在一个自由下降的电梯里，电梯里的人是感受不到万有引力的。

这是一个极端重要的发现，换句话说，在一个无引力场的空间，如果电梯向上加速的话，那么电梯里的人能感受到一个等效的引力。这就是等效原理。不过这个原理是很物理的，可以在黎曼微分几何里重新来说这件事情：弯曲流形上存在一个高斯法坐标系，使得流形上一点 p ，克里氏多夫符号函数在 p 点取值为0，度量在 p 点取为平坦度量。

爱因斯坦的等效原理和微分几何学里的高斯法坐标系是同一个事情。也就是说，在特定坐标系中，克里氏多夫符号函数会等于零。克里氏多夫符号函数和矢量沿着流形上的路径平行移动有关系。在欧几里得空间，也就是一般中学生学的几何中，背后有一个隐藏的假设，那就是：矢量在平行移动下是不变的。但弯曲的流形（引力场）没有那么好的对称性，矢量在平行移动的时候，移动后的结果是依赖于它走过的路径的。细节我们不再谈，反正，在很小的距离上，矢量平行移动的变化量和克里氏多夫符号函数成正比的。总之，矢量平行移动的结果是路径依赖的。打个比喻，这就好像蒋中正的曲线救国行动，行动的结果是依赖于他采取的曲线的。

爱因斯坦发现等效原理以后，问他的恩人格罗斯曼：“你觉得什么样的数学，可以描述电梯中失重的感觉？”

格罗斯曼沉默了半天，他是数学教授，所以说话颇为谨慎严谨：“也许高斯的微分几何学，可以帮你的忙.....不过，仅仅是也许.....”

听到这里，爱因斯坦觉得自己就像一个在黑暗中生长的蘑菇，已经快看到阳光了。他决定去学习微分几何学。

28 三年半的沉默

(1)

从1907年的12月到1911年6月（爱因斯坦定居到布拉格后几个月），爱因斯坦在引力问题上一直保持着沉默。

爱因斯坦在这个沉默期究竟在想什么呢？

但凡伟大的作品出世，一定需要有一定的时间来冷静和沉淀作者的思绪，所谓“文王拘而演周易，仲尼厄而作春秋；屈原放逐，乃赋离骚；左丘失明，厥有国语；孙子膑脚，兵法修列；不韦迁蜀，世传吕览”。大致也是一个意思，人生需要沉默期，大爆发之前需要积累。

爱因斯坦这段时间遇见了什么糟糕的事情了呢？

这段时间，他已经从专利局的技术职员混成了副教授，现在他来到布拉格大学，他要成为正教授了。这人生也是渐入佳境，爱因斯坦这个时候也做物理，但内心深处一直在思考一个问题，那就是引力场中光线弯曲的问题。

爱因斯坦的等效原理已经说明，牛顿的惯性力和马赫的引力是局部上不可区别的。这个时候，接下来的问题就是，如何用数学来描述引力。如果事情仅仅是牛顿引力，那牛顿的数学就是足够的。

但是，问题在于，牛顿的引力与狭义相对论是不相容的。

因为牛顿的引力的计算依赖于空间距离的计算，但是，在不同的参考系里，空间距离并不是一个不变量。一辆高速运动的火车，在地面参考系看来，长度会缩短，这是可以根据狭义相对论可以推导出来的所谓“尺缩效应”。

所以，必须把牛顿引力修改一下，使得它与自己的狭义相对论相互协调，这是爱因斯坦的初衷。大师一出手，必然是有大师的痕迹，所

以，在没有一定的把握之前，爱因斯坦选择了沉默。他补习了一段时间的微分几何。

因为朦胧的直觉是这样的，格罗斯曼告诉他，微分几何也许对他有益。格罗斯曼的这句话其实起到了决定性的作用。学微分几何没有几天，爱因斯坦就有一些朦胧的思想，这个思想就是：引力其实是几何学？

回顾一下历史。

牛顿说：存在虚拟的惯性力！

马赫说：其实是真实的引力引起了惯性！

爱因斯坦把两人的思想中庸了一下，说：惯性力等于引力。

到了这里，已经进入一个死扣，但爱因斯坦突然来了一招返璞归真，接着说：引力不存在，引力是时空的几何学。

这最后一句话，爱因斯坦将要大声地讲出来，这完全好像那孩子指出了皇帝的新装的秘密，新装并不存在，而引力也不存在。

为什么会这样？为什么要这样？真理是朴素的，也有的人说，数学真理就像是贝多芬的美妙乐章，不是人为雕琢出来的，而是不得不如此的作品，你改一个音符都不行。爱因斯坦的引力理论，也是不得不如此吗？如果真的非如此不可，那么显然是上帝假借了爱因斯坦之手在人间写方程了。历史是一步一步发展的，并且一定是具有雷同性和自相似的结构。

爱因斯坦为什么要用微分几何，这背后必然有天才的成分在里面，而微分几何刚出现的时候，也是经过天才之手。微分几何一开始，是高斯的杰作，高斯已经不满足研究一些曲线的弧长，他要研究曲面的曲率。高斯19岁的时候，他也和现在一般的高中生差不多，搞一些直尺和圆规来作图。他苦思冥想，就是要作一个正17边形。因为尺规作图的本质是在坐标平面上寻找具有特定形式的坐标，所以高斯从此对数论有了很高深的造诣。如果现在的高中学生也想有如此的成就，不妨找一找椭

圆曲线上的有理点。不过一般人才情总是难比高斯。他这番作完平面几何，就开始做球面，椭球面的面积，做完面积，他就做曲率。高斯对微分几何研究是有一些玩票的性质，他也是随便玩玩就可以了，得到了一个高斯绝妙定理和一个高斯-波涅定理以后就金盆洗手。接下来的事情由他的一个徒弟黎曼来出手解决。黎曼看着高斯绝妙定理出神，知道了一件大事情：几何对象的曲率的存在可以不依赖外部空间。换句通俗的话，高斯是用长焦镜头远距离来偷窥一个美女的身材曲率。而黎曼用的是X射线断层扫描技术来研究美女的身材曲率。高斯站在外面，黎曼在里面。

黎曼的几何学抛弃了坐标，黎曼看着高斯绝妙定理就已经晓得：几何曲率应该是与外部空间无关，甚至与坐标无关。

而爱因斯坦也领悟到了的一个道理：物理应该与坐标无关。

人心深处是相似的。

爱因斯坦在他的沉默期里开始与一些微分几何纠缠起来，他的蜜月期似乎又要来了。我们再来回顾一下这些思想史。狭义相对性原理说，“所有的惯性参考系中，物理规律是一样的。”基于狭义相对性原理和光速不变原理，爱因斯坦在1905年得到了狭义相对论。现在看来，狭义相对论是很自然的想法，但惯性系不是一个自然的概念。爱因斯坦不是一个普普通通的男人，他后来决定抛弃惯性系。在物理学里，惯性系是一个有特权的王国，爱因斯坦想，这个物理世界应该是民主的，不应该存在具有特权的参考系。他有了这样的思想——姑且称之为“参考系的民主”。为了研究万有引力，会发现万有引力的大小依赖于两个物体之间的空间间隔，但在四维几何里，三维空间间隔不是一个不变量，参考系改变以后，这个空间间隔就变化了，于是万有引力大小就变化，万有引力定律与狭义相对论的矛盾水火不容。这个矛盾大致可以这样看出来，两个物体之间的空间间隔依赖于观察者，所以在不同的惯性观察者看来，2个物体之间的万有引力大小依赖于观察者。这区别于库仑定

律，在库仑定律中，除了电力还有磁力，在电荷加速的时候还有辐射。万有引力定律对吗？狭义相对论对吗？爱因斯坦开始陷入了深深的思考。后来他意识到很重要的一点，那就是万有引力其实不是一种力，而仅仅是空间的弯曲的效应——也就是说，万有引力就好像是一个路灯下的影子而已，不是真实存在的东西。

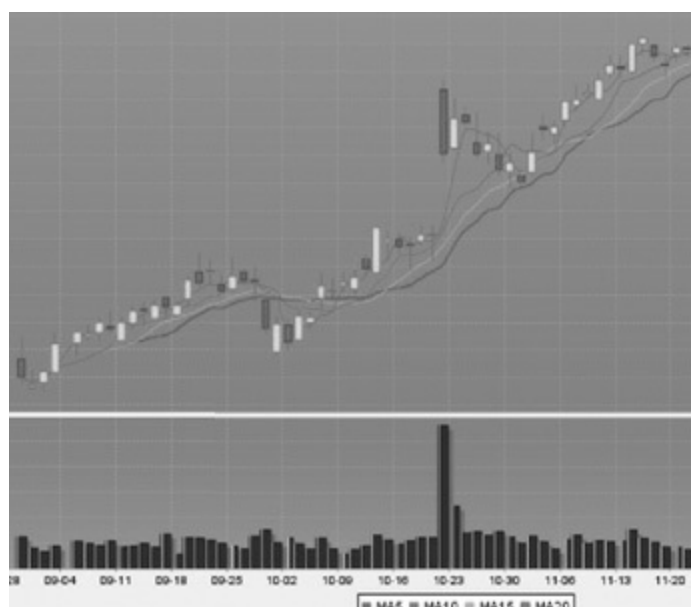
既然空间是可以弯曲的，那么他就可以使用微分几何这套数学工具了，也可以抛弃惯性系了，他于是抛弃了惯性系。惯性系成了一个弃妇，新人胜旧人，他也得到了新的原理。用现代语言来说，这个原理有三种等价的表达：

1. 广义相对性原理。
2. 广义协变性原理。
3. 微分同胚不变性原理。

在那时候，爱因斯坦提出了广义相对性原理：“所有的参考系中，物理规律是一样的。”有了这样一个原理，爱因斯坦要做的事情就是思考一下万有引力了，他要做的时候很简单，就是要让万有引力理论不依赖于参考系，不依赖于观察者。因为，爱因斯坦相信，物理规律是普适的，它是物理王国的法律，有上帝制定，对谁都一样，在任何时间任何地点，全是一样的，这正是人人平等的民主观念。就这样爱因斯坦用他的思辨构造了他的引力理论——广义相对论。在他的理论中，引力不是一种力，也许可以说，引力根本就不存在，所谓引力，其实是空间和时间（统称时空）被物质扭曲。正如一个人躺在席梦思床上把床睡得陷了下去。所以，空间上的点与点之间，其实是有人类的肉眼看不见的弹性纤维连接起来，这些弹性纤维组合起来就是一张弹性网络，这个网络的动力学规律就是爱因斯坦的广义相对论——正如一根弹簧的动力学规律是胡克定律一样。

引力为什么不存在？引力为什么是几何学？

实际上，引力场很像股票市场，如果你要了解股票市场，你会去看k线，一般来说，k线如果是分明的清晰的，也就是说，日k线压着周k线，周k线压着月k线，那么，这个股票一般被认为是处于上升通道。反正，如果k线相互绞着，看上去很紊乱不知所踪，那就是这个股票没有一个很明显的趋向，连主力都有点摸不着北。这是技术分析。



在股票市场中，资金是推动股价上升的基础，这就好像在广义相对论中，物质是引起时空弯曲的基础，但是，股价的上升会反过来因为资金获利回吐，在广义相对论中，时空的弯曲也会影响物质的运动。

对引力场，也是需要技术分析。因为引力场肉眼看不到，所以我们要看在引力场中运动的粒子，粒子的轨道就好像是k线。

所以，引力为什么是几何学？我们就需要看质点在引力场中的轨道，这个轨道会告诉我们一些信息。质点在引力场中运动，会在空间上画出一个曲线。这个曲线一般是不封闭的，因为空间曲线不是几何不变量，它是躺在空间上的，但是我们对时空的3+1分解是任意的（正如爱因斯坦转盘上的观察者们的空间感就和地面静止惯性观察者们的空间感

不一样），所以，这个空间曲线实际上是世界线在空间的投影。而在时空中的世界线才是几何不变量。我们说的质点在引力场中的轨道，就是在说世界线。

如果一个质点仅仅只受到引力的作用，那么，它在时空中的轨道就是测地线。大约爱因斯坦是从从这里开始类比起吧。

如果有很多这样的质点，可以发现在时空中会有很多这样的测地线。爱因斯坦在布拉格广场的时候，也开始有这样的朦胧想法。但他不懂得如何来描述测地线。

爱因斯坦站在办公室的窗边眺望远处的风景，玻璃窗太灰暗，在时空的困惑中。

他手上唯一有的是一个等效原理和一点微分几何的知识。他想，如果一个观察者趴在质点上随着质点在引力场中下落，那么，很显然，这个观察者感受不到引力的作用，所以，这个观察者在每一个瞬间都会认为质点是在走很短的一个直线。

也就是说，质点的加速度是零。

那么，牛顿第二定律可以写成这样的形式

$$m \frac{dx^2}{dt^2} = 0$$

这里，x坐标是自由下落观察者赋予的。这个x坐标系是什么呢？x和t分别表示空间坐标和时间坐标，但这个方程只在局部是成立的。

这个时候，爱因斯坦看到了另外一个微分几何里的方程，那就是测地线方程。

$$\frac{dx^2}{ds^2} + \Gamma \frac{dx}{ds} \frac{dx}{ds} = 0$$

爱因斯坦心里似乎有点谱了。因为这2个方程形式上蛮像的。在第2个方程中， Γ 是克里斯多夫符号。

第一个方程，是没有引力时候的运动方程（可以认为这个时候有惯性力）。

第二个方程，是微分几何里的测地线方程。

如果把s线长看成是固有时，那么，这两个方程大致上差了一个克里斯多夫符号，而克里斯多夫符号恰恰可以被看成是惯性力。

爱因斯坦有点懂了，原来牛顿第二定律和测地线方程，似乎是同一件事情，只要把克里斯多夫符号看成是惯性力.....

1911年，爱因斯坦在他的沉默期大致走到了这样的—个地方。里奇的微分几何，他是读了一遍又一遍。这也就是他三年半的沉默期一直在打磨的思想。

29 朗之万：双胞胎悖论

(1)

再说劳厄回到普朗克的身边，大大地夸奖了爱因斯坦一番，说爱因斯坦的狭义相对论真是在经典力学的瓦砾场上建立了一个华美的都城。普朗克也大为所动，于是写信给爱因斯坦讨论问题。讨论了半年，最后的通信让爱因斯坦看到以后心花怒放，因为普朗克说：“爱因斯坦是当代活着的哥白尼。”这是一句重话，相当于现代，杨振宁给一个大学生写信，说，你是当代活着的牛顿，这话传出去弄不好是要死人的。

历史的发展并不是一马平川的，而是蜿蜒曲折，狭义相对论的思想已经开始被普朗克等人在学术领域传播，这自然引起了一些观念上的革命，虽然爱因斯坦已经开始准备发展广义相对论，但狭义相对论的市场化进程却刚刚开始，消费者们还没有准备好接受这一款雷人的新产品。因为狭义相对论说，两个速度不一样的人，他们的衰老速度是不一样的。那还了得吗？

法国的物理学家朗之万也有点晕，他不知道到底有多少人已经理解了狭义相对论，于是他说：“全世界只有12个人能懂相对论。”朗之万的话一传出去，新闻界也听懂了，于是报纸上开始不断引用朗之万的这句诡异的话。这句话实在是太精辟了，看上去就像是一幅名画《最后的晚餐》，尤其是12个人，简直有了上帝的气息。而流言表明，这12个人，大多数是在柏林，而在法国的，显然也有一个人，这个人自然是巴黎的朗之万。可惜，朗之万是真的不懂狭义相对论，他不但不懂，而且还很糊涂，他的糊涂自然代表了时代的糊涂，因为他就像是科学界通往新闻界的喇叭，他又抛出了一个老妪能解的问题：双胞胎悖论。

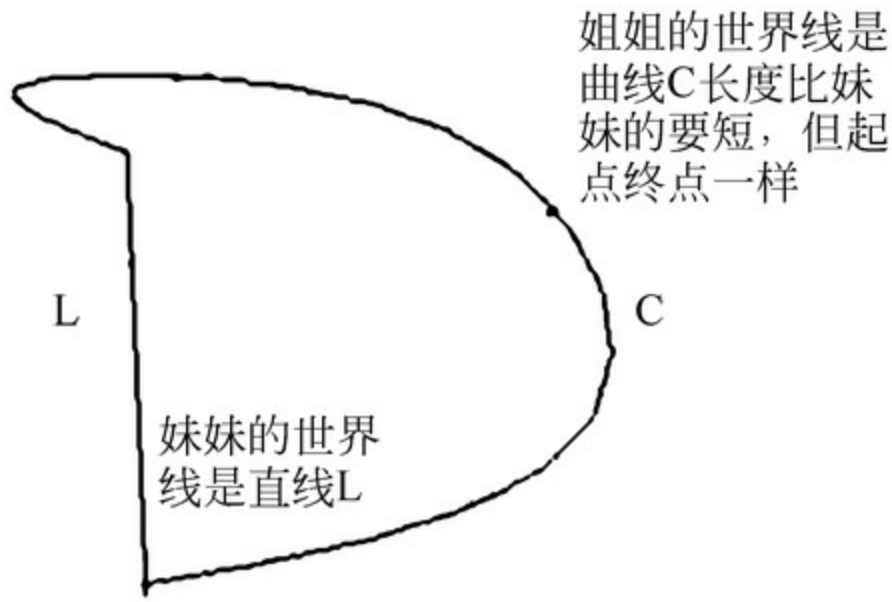
双胞胎悖论中的姐姐上天去了火星一番，妹妹留在地球，等姐姐回

到地球，发现自己还是如花似玉的大姑娘，但妹妹已经人老珠黄。悖论说，那既然运动是相对的，那为什么故事的结局不是姐姐人老珠黄妹妹如花似玉的版本？这个悖论一出来，街坊邻居们纷纷议论开了，茶楼酒肆咖啡屋里也争得甚嚣尘上像是被投了炸弹那样沸腾起来。狭义相对论被朗之万这样用捣糨糊的方法一弄，伤了大众的脑筋。如果读者们回望一百年来的狭义相对论发展之道路，会发现这个悖论一直是一座丰碑。这个丰碑的底座，镌刻着多少前仆后继夙兴夜寐的身影。多少人为了在这个丰碑上撞死而泪流满面痴心不改，多少人为了这个丰碑茶不思饭不想，多少人为了这个丰碑抛妻别子走上了民间科学家的漫漫野路？试想大地苍茫，岁月沉浮如波涛汹涌，又有多少人真正地理解了这个丰碑的意义？在1969年，苏联和中国在边界上有了军事摩擦，在珍宝岛上，双方边防军进行了小规模的战斗。在后来的谈判中，双方对到底是苏联士兵还是中国士兵首先开枪起了争执。当时在中国人们多数不相信爱因斯坦的狭义相对论，也觉得不理解。一些批判相对论的人认为，假如爱因斯坦的狭义相对论是正确的，那么在珍宝岛上到底是哪一方先发射第一枚子弹成了不可判断真相的政治事件——因为时间也是参考系依赖的。从狭义相对论的角度来考虑，1969年的珍宝岛，到底是哪一方先开枪？这个问题的答案依赖于在哪个参考系来看，是坦克司机的角度，还是克里姆林宫的角度，还是天安门的角度，还是月球的角度，还是火星的角度，还是织女星的角度……角度太多了，在这些不同的参考系下，哪一方先开枪，确实是各有可能。

今天，曾经困惑的灵魂还没有走远，但翻开历史我们可以看到吃人的历史上歪歪斜斜写着三个字：世界线。

从地球去火星一趟回地球的姐姐，她的世界线是闵氏时空中的曲线C。

在地球的妹妹的世界线是闵氏时空中的直线L。



姐姐的世界线长度比妹妹的要短，因此重逢的时候姐姐比妹妹年轻

世界线的长度 C 小于 L 。所以，姐姐的固有时间要比妹妹的固有时间流得慢。也就是说，姐姐是会比妹妹年轻的。这个问题所提出的时空背景流形是平坦的，所以这是一个狭义相对论问题。世界线的长短是一个积分过程，这个过程其实可以推广到弯曲的时空。世界线的长度是一个几何不变量，只有这样的不变量才代表真正的物理，也就是不会随着观察者本身的地位改变而改变。因为时间的相对性也出现在广义相对论中，而后者其实是描述万有引力的一门学问，所以世界线的长度这个概念继续有效。爱因斯坦后来成名以后，有一群人经常问他同一个问题：到底什么是相对论。爱因斯坦不胜其烦，于是这样给普罗大众介绍相对论，他说：“引力不是人们坠入爱河的原因所在。人世间，初恋是如此重要的生物现象，你怎么可能根据化学和物理学来解释呢？把手放在火炉上一分钟，你会觉得像一个小时那么久。而和你心仪的女孩偎依在一起，一个小时就像一分钟那么短暂。这就是相对论。”爱因斯坦这段话的本意大约是想说明，时间是私人的，也是相对的，并不存在绝对的时间。作者张轩中曾经多次在科普讲座中谈到双胞胎悖论的时候，应用了比较正统的相对论学家们的观念，首先参考的是梁灿彬老师的著

作《微分几何入门与广义相对论》——他是用世界线的长度代表固有时来解释这一悖论的，当然张轩中把四维加速运动看成是一个需要消耗能量的过程，总是说“谁消耗的能量与钱多，谁就更年轻！”。1966年，真的做了一次双生子旅游实验，用来判断到底哪个寿命长，同时也一劳永逸地结束了纯理论的争论。不过旅游的不是人，是 μ 子。旅途也不在天外，而是一个直径大约为十四米的圆环。 μ 子从一点出发沿着圆轨道运动再回到出发点，这同乙（姐姐）的旅行方式是一样的。实验的结果是，旅行后的 μ 子确实比未经旅行的同类年轻了。我们似乎可以这样作结论了：谁相对于整个宇宙作更多的变速运动，谁就会活得更长久。

(2)

可以说双胞胎悖论催生了一个相对论文化圈。接下来的问题就是要确定地球作为一个近似惯性参考系（或者说我们所处的时空近似为一个闵氏时空），它的近似程度是怎么样，不幸的是，这是一个技术问题，很难一针见血地加以说明，实际上说明这个问题，可能需要GPS的技术人才。

但可以借鉴马赫等人的说法，如果你抬头看星空，发现星星被没有像旋转木马一样绕着你的脑袋旋转，那么很有可能，你是处在一个惯性参考系之中。

你在地球上生活了那么多年，看到过星星都绕着你的脑袋旋转吗？

这在半哲学的程度上证明了，至少地球与宇宙中的很多星星保持了相对的平衡状态，地球与它们的地位接近相等，在整个宇宙背景中，地球是一个很普通的星体，它基本上可以被认为是一个近似的惯性参考系。

于是，对相对论文化圈来说，一个更难的问题马上就会浮现，到底什么是参考系？有没有让大家公认的良好定义呢。那么，首先地球到底

是不是一个完美的惯性参考系呢？我们要再次提到Sagnac，这个与朗之万与居里夫人同时代的法国物理学家。

Sagnac效应说的是，在一个旋转刚体上朝旋转方向和背着旋转方向发出的2束光再重逢的时候，它们花了不同的时间。因此，从物理的眼光来说，地球的自转使得地球其实是一个非惯性参考系。如果有工程师们读到这里，马上就可能分野为几种不同的派系，对这个实验的解读，也会有大不同，其中有一个派系，是否认此物理效应的存在，这个派系当然是最掩耳盗铃的一群人。因为从一个基本的实验事实可以看出：Sagnac效应是真实的物理。在GPS或者其他的各种全球定位系统中，都可以发现，在赤道上朝东和朝西2个方向发出的光，在被天空中的卫星接收到的时候，会有时间差，这个时间差的一部分，来自引力场的贡献，另外一部分，则来自Sagnac效应。

对于大部分人来说，最容易造成误读的，其实不是Sagnac效应本身，而是对于所适用的理论的选择。首先，我们可以肯定的是，Sagnac效应的计算，不需要使用广义相对论，这是一个纯粹的狭义相对论问题。地球作为一个小小的非惯性系是处在一个平坦的时空中的，它的自转会引起一点非惯性系的效果。在相对论的语言中，参考系是一组观察者的集合（也就是世界线的集合）。精确地说，地球作为一个非惯性参考系，其现实意义在于，地球上静止不动的人们，他们的世界线不是测地线，因为地球的自转，会引起这些人所对应的世界线随着地球自转一起在平坦时空中扭转——相对论工程师们大都吃过天津大麻花，也一定吃过油条，这种相互缠绕的结构，出现在世界线中，就是非惯性系。有的读者一定会更深入一步，说地球有引力场，它周围的时空一定不是平坦的。这当然也对，但对学习Sagnac效应的物理本质毫无裨益。美国航天局在2011年5月4日发布消息称，该局2004年发射的“引力探测器B”卫星已经证实了大质量物体的旋转还会拖动周围时空结构发生扭曲，这就“惯性系拖拽效应”。按照新闻报道中的科学家的通俗比喻，“惯性系

拖拽效应”有点像把一个橡皮球放入盛满糖浆的大碗，橡皮球或者说大质量物体的转动，会带动糖浆或者说时空结构跟着一起运动。“引力探测器B”的主要装备是4个超高精度的回转仪。当“引力探测器B”在距离地球约640公里的极地轨道上开始运转时，4个回转仪自转轴同时对准遥远恒星——IM Pegasi。如果地球引力不影响时间和空间，那么回转仪自转轴将一直指向初始方向。实际观测结果是，受地球引力拖拽，回转仪自转轴方向发生了可测量的细微偏移。但是，对Sagnac效应来说，“引力探测器B”所证实的“惯性系拖拽效应”只能说明地球是在自转，而且会引起可观测的物理效应，它与Sagnac效应没有一毛钱的关系。

麦克耳孙-莫雷实验是一个在地球上做的实验，实验的结果与时间无关，与空间无关，具有强大的说服力。光速不变原理已经成为狭义相对论大厦的基础，但对它的理解还是有很多相对论工程师觉得困惑。作者在经历了长时间的艰苦探索以后，发现这个问题与对光速的定义和理解有关。光速不变原理其实也正在被悄悄地误读，人们忘却了其本质的前提，而到处滥用此原理，错误地以为，在任何情景下，在任意定义下，超光速都不应该。但是，比如在宇宙学中，人们很容易计算出，远处宇宙空间的膨胀速度正在以比光速还要大很多的速度发生。在时空中，如果单纯使用光子运动的距离，除以光子运动的时间，得到的也一定是一个模糊不清的概念，众所周知的是，在相对论中，距离和时间，都不是固定的量。因此如果可以重新来描述此光速不变原理，应该说：“在任何观察者对物体的当时当地的局域测量中，该物体的速度都不能超越光速，而光速是一个恒定的常数。”

因此，严格地指定是什么观察者在观测，是谈话和讨论的基础。

在相对论中，有两种不同的测量方式。这两种方式如下：

1. 依赖光信号的传播，只有一个观察者来进行测量。
2. 不依赖光信号的传播，在时空中存在各种观察者，在当时当地进行直接测量。

要搞清楚此两种基本的测量方式，是很重要的。光速不变原理，符合第二种测量方式。那么在Sagnac效应，到底是谁在测量呢？转盘上的观察们到底看到了什么？

麦克耳孙-莫雷实验是一个非常精彩的物理学实验，如果大家去看陈应天老师的科普书《相对论时空》，就可以看到那里有一张漫画，画的是怎么测量一架飞机自身的飞行速度，可以通过一个声纳朝2个相互垂直的方向发出声波，然后通过接受到这两个回声的时间差来计算出飞机的速度，这是一个很巧妙的发明。

麦克耳孙-莫雷实验也是同样的原理，但是，结果却与飞机的情况大相径庭。

在两个垂直的方向上发射的光，人们本来渴望能够测出地球的公转速度，但是，现在却是测量不出来的，地球就好像一架飞机，但无法用现在光传播的时间差来倒推出地球的速度。换句话说，在这个情景下，光的传播在两个方向上没有表现出时间差。

什么是时间差？这里的时间，其实是一个观察者所使用的时间。

这个观察者，也可以被认为是站在干涉仪之上的。

可是，这个干涉仪明明是在地球上，随着地球在自转，为什么测量不出时间差来呢？Sagnac效应不是告诉我们，在这样的光路中，应该存在时间差吗？为什么没有干涉条纹呢？地球的自转如何体现？

其实原因很简单，从理论上来说，地球的自转应该会给麦克耳孙-莫雷实验带进Sagnac效应。但是，这个效应太弱了，麦克耳孙-莫雷实验测量不出来这个微弱的干涉效应，把它当作噪声忽视了！

自转的地球上麦克耳孙-莫雷实验中Sagnac效应太弱的原因很简单，其中一个主要的原因是地球自转的角速度太小，地球24个小时才自转了 360° ，而一般用在Sagnac效应上的旋转平台，每分钟都能转几百转，因此不可同日而语。

这件事情其实告诉我们，物理是一个统一的整体，各种物理效应都

可能出现在同一个实验中，但我们要抓住实验的最本质内涵。麦克耳孙-莫雷实验的本质是说，在忽视地球引力场和地球自转的影响下，地球本身是一个惯性系，在惯性系中，光速是不变的。

如果翻开张元仲老师的《狭义相对论实验基础》，在第13页我们可以读到：“由于地球每秒以三十公里的速度绕太阳公转，因此在一年四季的不同时间内，地球上的实验室就是不同的惯性参考系……”

以Sagnac效应作为基本原理的光纤陀螺仪器在导航上大量使用，但对Sagnac效应的相对论解释却是大多数工程师们所不具备的。如果把光纤陀螺仪放在地球上，如果光纤陀螺与地球的自转平面垂直，那么你看不到干涉条纹，但是，如果光纤陀螺与地球的自转平面平行，那么你可以看到干涉条纹。这个基本的实验事实似乎说明：地球的自转可以被测量出来。

读费保俊老师的书《相对论在现代导航中的应用》，就会发现Sagnac效应已经清楚了。总的来说，这个是一个起源于度量之中出现时空交叉项而引起的力学效应。如果翻开《相对论在现代导航中的应用》此书的第103页，我们就可以看到，在地球系中，因为考虑到地球的自转，在地球系中光传播的运动方程可以被解出，从地球的赤道面上同一点发出的两束光，朝东发出回到原点的时间，比朝西发出回到原点的时间，要多花207纳秒。

我们已经知道，每秒钟光可以绕地球的赤道走大概7圈半，因此，这207纳秒也不算是一个很小的量，尤其在精确制导的导弹上，这207纳秒可以引起导弹几百米的误差。如果不消除这个Sagnac效应的影响，本来要打到长江南岸的导弹，会直接飞到长江北岸。

因此，可以通过Sagnac效应来大致估计一下地球接近于一个惯性参考系的程度，这纯粹是一个工程学上的问题。

萤火虫会在黑夜里留下一个光痕。看到一个空间轨道，如果你总能想到沿着时间在时空中拉成一个世界线，于是，一些基本的问题会像海浪一样涌上来。你还会问：

1. 存在不存在闭合的世界线？
2. 时空流形上的两个时空点之间，可能不可能有无穷多条测地线连接？

这样的怪问题会越来越多，但只有这样，才可能有异常的眼神，人们看到的是时空点的集合，相对论专家看到的是线条（世界线）的集合，这样下去，才能解决朗之万与Sagnac等法国物理学家们心头的困惑。

30 广义相对论

(1)

爱因斯坦在伯尔尼大学做无薪的讲师，觉得自己地位颇低，虽然普朗克把自己评价为活着的哥白尼，但江湖上还是不太承认自己的独特地位，这让爱因斯坦内心有一种苦闷，讲师的地位实在是太低了，爱因斯坦于是回到了自己大学时代度过的那个城市——苏黎士，苏黎士风景秀美，苏黎士大学可以给他副教授的职位。

爱因斯坦觉得，能再次来到苏黎士，天地开始开阔起来。

这个时候，狭义相对论已经开始在市场上缓慢流行，爱因斯坦不会停止自己的脚步，他脑子里有一个巨大的问号，这个问题在牛顿的《光学》一书里也曾经提出来过，那就是：

引力场的存在会不会使得光线走过的空间路径弯曲？就好像斜抛出一个苹果，苹果在空间将划出一条抛物线？

爱因斯坦知道有一个关于光线弯曲的重要定理，这个定理就是菲涅尔定理，描述光的折射。

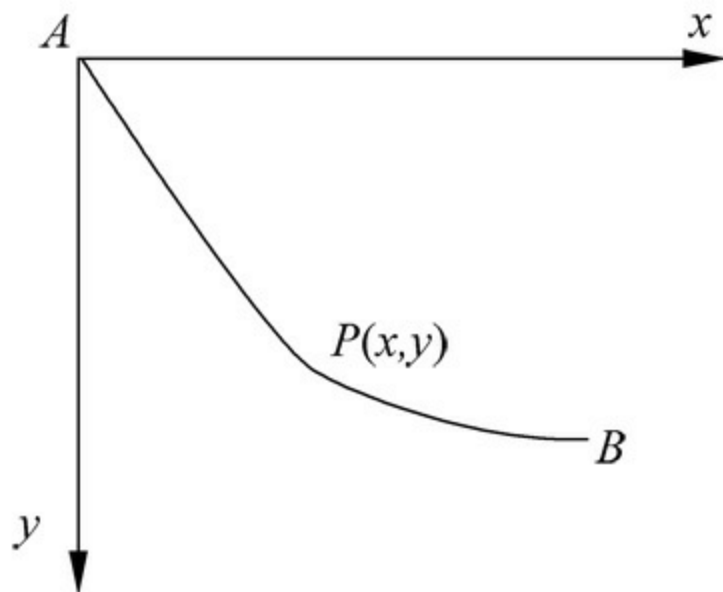
1. 光在同一个媒介沿着直线传播；
2. 在不同媒介的交接面上，光线发生折射，入射角和折射角的正弦之比是折射率的倒数比。

这个定理是非常强大的，但它本质上是费马原理。光从a点到b点的传播时间取极小值（或者说光程取到极值）。用这个原理可以解决一系列平面几何问题，比如可以证明在任意三角形的内接三角形中（每个顶点在依次在外面那个三角形的边上），以垂足三角形的周长最小。

爱因斯坦在考虑光线弯曲的过程中想到过这个原理，他发现在理论中存在以下雷同的地方：

1. 最速降线
2. 光线折射

“最速降线”的问题是一传统的问题。如果在平面重力场中，高处有A点，低处有B点，如果A，B不在一条铅直线上，那么，在两点之间连一条曲线，问什么曲线能让小球沿这个轨道滑下来用的时间最短。



函数 $y(x)$ 的曲线

牛顿当然思考这样的问题，但不知道怎么做。问题留给了伯努利家族。伯努利兄弟自然是技压群雄，解答了这个问题。

可以写出这个时间的积分。

$$t = \int \frac{ds}{v}$$

其中， ds 是曲线的弧长， v 是速率。

这个积分写在直角坐标系中，根据能量守恒，一定是很容易写的。

问题的关键是，你要求 t 最小，但曲线的形状 $y(x)$ 没有确定，所以这个积分实际上是一个泛函（注： t 是函数 $y(x)$ 的函数）。最速降线的 t 是路径 $y(x)$ 的一个函数。

这一点是非常重要的。——如果读者们有宏大的视野，可以相信，这个问题可以用光线在一个变化折射率的介质中的传播时间最短来模拟。但无论这个问题的模型是什么，总之，这是一个欧拉-拉格朗日变分问题。

好了，以上这个积分其实可以看成是一个阿贝尔变换。

在第2章讲过，阿贝尔是挪威的青年，他的生命短暂，但万古长存，他死后就算过2000年，只要还有人类，依然会有人怀念他。他曾经也考虑过引力场中的路径问题，不过换了一个版本。

阿贝尔的问题是：如果有一座山，一个小球因为重力从山上滚下来的时间 T 是山的高度 h 的函数 $T(h)$ 。 $T(h)$ 的表达式就是上面我们讲的积分，如果你已经知道 $T(h)$ ，那么你能不能反推出这个山的形状。

答案是肯定的，这就是阿贝尔变换存在反变换。正如傅里叶变换存在反变换一样的。

阿贝尔死于肺结核，年轻的阿贝尔死的那个晚上很是孤独。爱因斯坦还活着，爱因斯坦并不清楚阿贝尔的故事，也不知道监狱里的那个发明李群的索菲斯。

有很多东西是爱因斯坦不知道的。

最速降线需要的时间 t 是一个积分。要求选择积分的路径，让 t 最小。在这里 V 是一个常数（总能量为1）减去高度（重力势能）开根号，也就是动能开根号 $V = \sqrt{1 - y}$ 。

而因为光线在运动过程中也走了最小的时间。

$$t = \int \frac{ds}{V}$$

在平面变折射率的介质中， V 是坐标的函数。

$$V(x, y) = c/n(x, y)$$

显然，如果取折射率满足 $n(x,y) = \frac{1}{\sqrt{1-y}}$ ，那么，光线在变折射率中的弯曲的轨道就是最速降线的轨道。

换句话说，几何光线的弯曲可以用重力场中的有质量粒子的下落轨道来模拟。

这在数学意义上把光线弯曲和重力联系了起来。这就是山寨版本的广义相对论，也是任何一个具有初中物理水平的人可以作出来的。爱因斯坦在1912年，开始思考引力场中光线弯曲的问题，也采取了类似的办法，但结果不完全对，为什么呢？因为引力场实际上不能用一个标量势函数来描述，而是一个4×4的矩阵。在某种意义上，我们也可以称爱因斯坦的广义相对论是一种“矩阵场论”。

(2)

爱因斯坦的狭义相对论出来以后，广义相对论也不是没有苗头的。因为埃伦费斯特马上提出了一个转盘的悖论：根据爱因斯坦的狭义相对论，运动速度快的物体长度要收缩。所以，如果一个旋转的圆盘，它的角速度是一样的，但线速度随着半径不同会不一样，于是，在一个宏观的惯性系的观察者们看来，这个圆盘的外围收缩得比内部要厉害一些。

这样的话，整个圆盘就要被扭曲了！！！！

可是，跟着爱因斯坦转盘一起旋转的观察者看来，转盘是静止的，所以转盘根本不会扭曲。

这就是爱因斯坦转盘引起的一个深刻的问题。

牛顿的旋转水桶和爱因斯坦的转盘，在物理模型上来说是一样的。但细节上有所区别。

我们看到水桶的表面弯曲起来，在牛顿的语境中，这和狭义相对论是没有关系的，本质上，这样的水面是一个等势面，也就是引力势和离心势的和是常数的一个面。

我们说爱因斯坦的转盘也会扭曲，这是因为我们考虑了狭义相对论的尺缩效应，并且这个结果是一群静止的惯性观察者们看到的。它们能测量转盘的周长和速度，会发现圆盘转起来以后有了扭曲。可是，跟着爱因斯坦转盘一起旋转的观察者们看来，转盘是静止的，所以转盘根本不会扭曲。所以，这里就存在这样一个问题，爱因斯坦转盘到底有没有扭曲呢？

要回答这个问题，必须搞清楚一件事情，那就是，那一群静止惯性观察者们和另外那一群随转盘一起旋转的观察者们，它们之间到底是什么关系？爱因斯坦转盘上的观察者们，是趴在转盘上的一群看客，从数学上来说，他们就是平坦的闵氏空间上的一个类时矢量场——他们的世界线是这个矢量场的积分曲线。这个矢量场的一个显著的特点是——它不是超曲面正交的——因此它们根本不存在统一的空间观念。

在广义相对论中，一个矢量场就是在时空上的每一个点指定一个方向，就好像在纸面上画的电场线是一样的。但有的时候，相对论学家把矢量场和它对应的一形式场看成同一个事物。一形式场（one form）是非常重要的，因为在数学上，有一个很重要的微分几何引理，被称为庞加莱引理，这个庞加莱我们提到过的，是一个数学奇才。但这个引理可能是张冠李戴放在他的名下，实际发现者另有他人。庞加莱引理是很广泛的，基本上包含了矢量分析的全部内容。应用到一形式场，如果一个一形式场 p 满足 $dp=0$ ，那么，可以知道，局部地有 $p=df$ 。这里 f 是一个函数，也就是说， p 这个一形式场可以看成是一个函数 f 的微分。

这个时候，我们说， p 是可积的。

换句话说，判断 p 是不是可积，我们要看 dp 是不是等于0。

已经说了，一形式场 p 对应的就是矢量场 V ，如果 p 是可积的，那就等于说，与 p 对应那个矢量场 V 就是可积的。这个时候，直观一些地想像，就是，你可以找到一个超曲面，使得这个超曲面的法方向就是 V 。

但这样的情况很少发生，读者们需要注意的就是，可积的矢量场是

非常稀少的。

并且，我们还需要知道，对于类时的矢量场来说，可积性这个条件是比较超曲面正交的判断条件要强一些。换句话说，可积的，那一定是超曲面正交的，但反过来不一定对。

超曲面正交也有自己的判断公式，这被称为FROBENIUS定理。简单地说，如果一个一形式场 p 是超曲面的、正交的，那么，我们可以知道必然有 p 和 dp 的外积是等于0的。

在数学上来说，超曲面的正交相当于说 $p=gdf$ ，其中 g 和 f 都是函数。

或换句直观的话说，这个时候相当于说 p 和 dp 是“平行的”。

超曲面正交的意思是说，与矢量场 V 垂直的矢量们是可积的。对于爱因斯坦转盘观察者们来说，与这些类时的矢量场垂直的矢量场们如果可积，那正好是一个空间部分。但不幸的是，爱因斯坦转盘观察者们作为一个类时矢量场不是超曲面正交的，所以对他们来说，连空间的定义也有点问题了。

科普读者们需要知道，空间是依赖于观察者的，不同的观察者有不同的空间。

(3)

1915年6、7月，爱因斯坦在哥廷根作了6次关于广义相对论的学术报告，主要是和希尔伯特讨论场方程该如何从作用量里导出。11月爱因斯坦提出广义相对论引力方程的完整形式，并且成功地解释了水星近日点进动。爱因斯坦是从广义协变性里推导出引力场方程的。而到了1916年，3月他完成总结性论文《广义相对论的基础》，广义相对论正式出炉了！几乎在同时，数学家希尔伯特构造了引力场的希尔伯特-爱因斯坦作用量，用物理学的标准化程序通过对度量变分也得到了引力场方

程，他说：“哥廷根大街的每一个小孩都比爱因斯坦更懂四维几何，但发明广义相对论的是爱因斯坦而不是数学家。”

对于物理学家来说，作用量是物理中最基本的，有了作用量物理学家就有了一切，因为上帝在创造这个宇宙的时候，总是有一个单纯的原则，那就是让作用量取到极小值。

无论怎么样，爱因斯坦方程是完全不依赖坐标系而写出的张量等式，是天人合一的典范，它的出世，表明纯粹理性具有非凡美感，人类心智，极富荣耀。

$$G_{ab} = T_{ab}$$

也就是说

$$R_{ab} - \frac{1}{2}Rg_{ab} = T_{ab}$$

其中 R_{ab} 是里奇张量， T_{ab} 是物质场的能量动量张量。可以在数学上检查下方程的左边求协变导数等于零了。

这里要注意的是， T_{ab} 是非引力场的物质场的能量动量张量，引力场的能量动量怎样描述，从开始到现在，都是让人困惑而痛苦的。后来邦迪和塞司有一些关于引力波携带的能量的工作，才说明这一点非常复杂。爱因斯坦一开始写出的方程与正确的方程是有点神似的，但物理上却过不去，因为方程的右边，物质场的能量动量张量的协变导数必须是零——能量动量对任何参考系都应该是守恒的。但是，如果要把协变导数作用到方程的左边，却一般不是零，除非你对曲率提出额外的要求——要求标量曲率是一个常数。换句话说，左边的曲率，其上面天然的限制条件是毕安基恒等式，但这个数学条件与能量动量对任何参考系都应该是守恒的物理规律在爱因斯坦最初的方程中是冲突的。

所以，爱因斯坦知道，需要对方程做一些微调。（微调这个词语就好像中央银行说要微调货币政策，往往会引起股票市场的大变化，因此

爱因斯坦对方程的微调也有很深刻的影响。) 微调以后的方程才是正确的爱因斯坦引力场方程。

爱因斯坦的引力方程显得有点来历不明，被认为是神来之笔，后来爱因斯坦的笔记本被发现，大家才发现原来这个笔记上也是一片潦草、苦闷和彷徨。

爱因斯坦引力场方程出来了，这方程显示，几何等于非引力物质场。这是很奇怪的一个方程，那么，引力还是一个物质场吗？如果没有非引力物质，那么， $T_{ab}=0$ 。

这个时候，叫做真空引力场。

外一篇 真空爱因斯坦引力场

在相对论专家彭罗斯的科普写法中：

$$\text{黎曼} = \text{里奇} + \text{外尔}$$

真空的引力场是非常有趣的，因为真空中，黎曼曲率只有外尔张量，所以，后来有人开始研究真空引力场的代数分类，或者说研究外尔张量的代数性质，就好像研究矩阵的特征值一样，这个故事是相当有趣的。这代数分类，称为佩多夫（Petrov）分类。

佩多夫是苏联人，他在1954年左右开始考虑外尔张量或者黎曼张量的代数分类，到1966年，思路已经完全成熟。他显然是苏联人中研究相对论而在历史留名的少数人了。当然其他的俄国人就是朗道、泽尔多维奇等人，朗道写的《经典场论》被认为是一代经典。朗道研究相对论的时候，有个中国人跟他一起做研究，他就是段一士。段一士被朗道认为是无比聪明的中国青年。这是几十年前的事情了。段一士对广义相对论的能量问题，有一个自己的表述，被称为“段一士能量表述”。佩多夫也是最早几个认识到20世纪20年代伯克霍夫的定理有缺陷的人之一。他在1963年指出这个错误，离伯克霍夫证明那个定理已经40年了，而伯克霍

夫1944年就去世了，他活着的时候未能看到自己的错误被指出，不失为了一件快慰的事情。

什么是外尔张量 W_{abcd} 的代数分类呢？佩多夫用的是线性代数的方法，因为外尔张量的下指标 (ab) 和 (cd) 是对称的，它可以被看成是一个对称矩阵。

给定一个矩阵 M_{ab} ，再给定矢量空间的基，那当然可以把这个矩阵写出来。这个矩阵无论怎么复杂，总可以讨论它的本征矢量。当然本征矢量很有可能是重复的，也可能找不到它的本征矢量。

对于外尔张量 W_{abcd} ，情景很类似，这个时候，佩多夫只考虑它的类光本征矢量。当然这四个类光本征矢量也有可能是有重复的，或者找不到这样的类光本征矢量。以下的数字 i 表示 i 次重复的本征矢量。

(1, 1, 1, 1)

(2, 1, 1)

(3, 1)

(2, 2)

(4)

(退化)

以上五种情景就是外尔张量的分类。对组合数学熟悉的人也许会惊讶，这不正是整数4的无序分拆吗？ $4=1+1+1+1=2+1+1=3+1=2+2=4$ 。这些型号的名字分别是第一类叫I型，最后一类叫O型——外尔平坦， $(2, 2)$ 型叫做D型。史瓦西时空和克尔时空全是D型时空。

有了对外尔张量的分类的数学，人们才能很好地处理引力辐射问题。莎斯（Sachs）得到了无质量场的剥皮（peeling off）定理。后来彭罗斯则用旋量语言很简单地重新得到了皮特夫分类。外尔张量其实对应一个前面说过的自旋为2的旋量场。任何一个自旋为 n 的无质量场全可以用 $2n$ 个2分量旋量的对称直积来表示。在每一个时空点，这 $2n$ 个2分量旋量——如果你还记得“旋量是类光矢量开根号”——对应 $2n$ 个类光矢量，

这些类光矢量被称为这个自旋为 n 的无质量场的“主类光方向”（principal null direction）。这些类光矢量在任一时空点的光锥之上。如果这些主类光方向全部重合，那么这个场就是类光的。对于外尔张量，是自旋为2的场，它有4个主类光方向，皮特夫分类说明了这4个主类光方向的重合情况。

如果时空是里奇平坦的，那么它可能是代数特殊的。一个真空引力场称为代数特殊的，即外尔张量不是I型或O型的，或者说外尔张量的主类光方向有重合，那就是代数特殊的。

哥腾伯格-塞司（Goldberg-sachs）定理说的是非O型真空引力场是代数特殊的充要条件为它的主类光方向所定出的类光测地线是无剪切的（shear-free）的。剪切是矢量场的形变（distortion）的对称无迹部分。扭转是矢量场的形变的反对称部分。一个扭转的但没有剪切的类光测地线汇被称为“罗宾逊线汇”，罗宾逊在这里也到达了发现扭量理论的边缘。

总之，这样就把时空几何和类光测地线在某个地方联系了起来。

相对论作为一个时空的理论，其基本的意思就是说，时间和空间是依赖于观察者的——但时空是绝对的，在爱因斯坦转盘这个模型里，时空还是平坦的闵氏时空。

在广义相对论中，矢量场也可能会有结构。比如光线是可以被看成是类光矢量的积分曲线的，所以，如果你看到一束光线穿过一个凸透镜，你会发现光线在空间上会被汇聚到一个焦点之上。引力场也有同样的效果。

在这个意义上，类光测地线在时空的几何分析中会有极端重要的作用。

凸透镜成像在任何照相机镜头中都有应用，很好的镜头是需要做像差分析的。相对论中也是如此。

31 美国空军的研究员

(1)

在前面第30章，我们的相对论之旅基本已经走到了终点，接下来的历史不应该被称为历史，因为很多当事人还活着。如果有一天，你可以像钱德拉塞卡一样，坐在从印度到英伦的邮轮上，在夕阳的红晕之下远行，带上一本钱德拉塞卡的书，在漫漫旅途里欣赏海鸥在天边飞行，你也许会惊叹，世界竟然有如此美妙的风景，你也许同时会惊叹，每一只海鸥是死去水手的灵魂。翻开钱德拉塞卡的书，《黑洞的数学》前面的一页里印着两张照片。其中一张照片是史瓦西的，另外一张就是克尔的。如果问1963年以后的相对论工作者，从1916年广义相对论诞生到1963年，最激动人心的事件是什么？答案当然是没有什么激动人心的事情发生，除了克尔解的发现。

1963年有一个相对论专家和天体物理学家的交流会，这个交流会共七天，每天从早上8:30到凌晨2:00，来自美国空军一个研究室的研究员，他在那里做了一个10分钟的演讲，他一上台，天文学家和天体物理学家们就没剩几个，剩下的，也都在小声讨论自己的话题，还有的就是在打瞌睡。

这个研究员就是克尔。他当时是美国空军的研究员。

克尔发现了一个爱因斯坦引力场方程的解——就好像当年塔塔里亚发现了三次方程的解一样，他的内心也是相当地冲动——他找到这个解可以说使用了佩多夫等人的结论也用到了哥腾伯格-塞司（Goldberg-sachs）定理。此解描述了黑洞作为一个定向陀螺如何带动周围的时空旋转。这的确是相对论历史上少有的真正意义上的进步。克尔时空有一个特点，那就是旋转星球的外部时空，不可避免地会被星体所拉动，这看上

去，非常像一个旋涡。

因为大多数的星体总是在转动，于是就有角动量，这样的时空，假如转动不能忽视的话，那么它的解显然不能用史瓦西解来研究。所以，克尔解的现实意义是巨大的。

在这个解里，因为存在着相互对易的2个凯林矢量场，这2个矢量场是等度量群的生成元。所以说，它的等度量群是一个阿贝尔李群。

自由落体运动是伽利略最喜欢的，在相对论中，它同样受到所有人的青睐。当考虑在克尔时空外部有一个粒子作自由落体运动，也就是走测地线的时候，我们如何来解出这个测地线。我们需要运动常数来列出方程。对于克尔时空来讲，最重要的性质之一是在它那里存在一个凯林张量。这个凯林张量是一个 $(0, 2)$ 型的张量。它的存在，将使得克尔时空中的自由下落粒子，沿着这个粒子的世界线看，它的能量是运动常数，它的角动量是运动常数，它的质量是运动常数，还有一个运动常数就是这个凯林张量场与粒子的四速的平方相互缩并得到的，彭罗斯称之为卡特常数。卡特是相对论研究中的一个著名人物，他和罗宾逊等人研究克尔黑洞，非常入迷，后来他们发现了比较著名的黑洞无毛定理：渐近平坦的稳态黑洞必然是克尔-纽曼黑洞。也就是说，稳态的黑洞，只可能有三个自由参数，一个是质量，一个是电荷，一个是角动量。罗宾逊是英国的绅士，他和卡特的工作无非是一系列数学，比如说轴对称黑洞必然是稳态的，但反过来可以问的是：稳态黑洞是不是一定是轴对称的。

具有电荷的旋转黑洞非常像一个质子，但它必然比质子大，因为形成黑洞有一个质量下限，那就是以美国原子弹之父奥本海默的名字命名的质量下限，大概是2~3个太阳质量。也就是说，在经典广义相对论中，不能把质子想像成为一个黑洞，因为两者在尺度上，具有天壤之别。但带有电荷的旋转黑洞是最普遍的，这就是克尔-纽曼黑洞。

克尔-纽曼黑洞是最普遍的黑洞，在星体旋转的时候，它的外部不

是真空的，而是有电磁场。这个时候，叫做电磁真空，时空的标量曲率是零，但时空的黎曼曲率不是零。克尔-纽曼时空的标量曲率为零，原因当然是因为电磁场是光子场，光子是零质量的。所以无质量场的能动张量总是没有迹的。克尔-纽曼黑洞具有电荷和质量，但这里面还有无穷的奥秘，一个很奇怪的事情出现了，假如这个黑洞的电荷或者角动量远远大于它的质量，那么黑洞的奇点就要裸露出来。假如真的存在这样的黑洞，那么对于人类来说是很危险的，因果性被很严重地破坏了。因为裸露的奇点要吞噬周围的物质，是没有预兆的。这有点像热带丛林沼泽里的鳄鱼，突兀地吞噬前来戏水的羚羊。单说电荷，在直观上，我们似乎没有能力要求一个黑洞的电荷不能远远大于它的质量。这就是一个荷质比的问题，对于基本粒子，我们可以用磁场来研究运动的带电粒子的荷质比，可以想象，一个粒子的荷质比越大，它在磁场中走的圆周的曲率也就越大。这个是简单的高中物理就可以解释的，也就是洛伦兹力提供粒子的向心力，那么荷质比正比于圆周的曲率。但对于黑洞来说，这个荷质比似乎要满足一定的限制，你不能太大，太大了，黑洞的奇点就裸露出来了。

裸奇点看来是一件糟糕的事情，因为这样的奇点可以突然吞噬黑洞外的事物。设想一下，你正在看的一本书，因为这本书撞上了裸奇点后突然消失。这似乎很诡异，但裸的奇点，它的外面没有视界包裹，像一个没有穿衣服的裸女，让相对论专家对她既爱又怕。彭罗斯猜测，裸露的奇点不应该在自然界出现。这就是著名的宇宙监督假设：上帝禁止裸奇点。

对于微分几何学家来说，彭罗斯的猜测无非是一组艰深的微分方程。谁能够解决这个问题，谁就能够名垂青史。但彭罗斯本人也无法解决这个问题，所以这无疑是一个非常困难的题目。

“大质量恒星坍塌后形成黑洞”、“黑洞分为静止黑洞和旋转黑洞，旋转的黑洞就是‘克尔黑洞’”。2011年8月8日下午，国际著名天体物理学家、数学家、物理学家罗伊帕特·里特·克尔（Roy Patrick Kerr）教授，在云南大学的科学大讲坛上讲授《黑洞和时间》，在近一个半小时时间里，克尔教授和场内观众一起就“黑洞”理论进行了知识普及和问答互动。这可能是当年的这位美国空军的研究员第一次来到中国访问讲学，当时的媒体对他的来访基本没有什么报道，可见他与霍金相比，在公众的知名度犹如云泥之别。

那天的云南大学科学馆因有克尔教授的到来蓬荜生辉，来自社会各界的学者、专家、学生和天文爱好者齐聚于此，只为听取这位为人类天体物理学贡献了卓越贡献的老教授的演说，可容纳近千人的会堂座无虚席，连过道和门口都站满了人。演说开始前，在座的每名观众都获得了一个桂圆——因为旋转黑洞的结构很像一个桂圆，克尔教授的演说就从这个桂圆开始，讲述了他28岁时如何求解出爱因斯坦引力场方程的一组精确解，随后他为大家介绍了20世纪60年代天文学和物理学面临的问题、黑洞的科普知识以及“时间”概念需要怎么改变等问题。

克尔教授毕生从事广义相对论、引力论研究，因成功解决旋转黑洞（即克尔黑洞）的引力场和时空问题而为人知。

报告会后的现场互动也你来我往，相当火热。在场听众所提问题逐一涉及“反物质”、“平行世界”、“白洞”、“虫洞”甚至“时光倒流”等诸多概念，都被克尔教授一一解答。这个时候克尔已经到了古稀之年，岁月如一把杀猪刀，时间在他脸上刻满了皱纹，几乎没有人知道他年轻的时候曾经在美国空军当过研究员。

32 贝肯斯坦和霍金

(1)

贝肯斯坦出名的时候还是一个年轻的研究生，那是在1972年，他20多岁。那时候霍金也还很年轻，30岁，但霍金已经很有名气了，因为1969年彭罗斯证明了第一个奇性定理之后，霍金迅速地跑上去证明了第二个奇性定理，1970年霍金和彭罗斯（R.Penrose）合作，证明了宇宙奇性定理：在极一般的条件下，按照广义相对论，宇宙大爆炸必然从一个奇异点开始。由此，他们共同获得1988年的沃尔夫物理奖。

霍金，1959年17岁的时候考入牛津大学学习物理，1962年秋天他到剑桥读研究生，开始研究广义相对论和宇宙学，1965年获博士学位。

而贝肯斯坦的故事，与一种叫“熵”的东西有关系——“熵”简单地说就是一个物体的混乱程度，如果一个女明星的私生活很混乱，我们可以说她的私生活熵很高，熵总是会越来越高，这是物理规律。

“熵”这个字非常漂亮，秀才读半边，一般人就是不认识也可能知道它的发音，“商”，——猜想它与除法有关系。但从字面上看，它与“火”有关系，或者说与温度有关系。没有错，熵S是能量U与温度T的商。

$$S=U / T$$

在刘慈欣的科幻小说《三体》中，也有“低熵体”的说法，智慧生命要维持自己的低熵状态，必须把高熵排泄给外部环境。如果读者想要计算一个人一天要排出多少熵到环境里，可以参考赵凯华老师的书《定性与半定量物理学》。

贝肯斯坦提出黑洞熵的时候在美国普林斯顿大学，是对美国政府无比忠诚的物理学家惠勒教授的博士生。贝肯斯坦和霍金一样，出名很

早，算是一个典范。他年轻的时候天高云淡，历史就给了他机会。他1972年关于黑洞熵的研究直到今天还是量子引力中最为重要的工作。贝肯斯坦的黑洞熵公式里出现了 h ，出现了 G ，出现了 c ，这说明黑洞熵与量子力学有关系，也与相对论有关系，于是，就与量子引力有关系，从道理上来说，这是现代人类文明的最高成就了。

那是在光辉的1972年。

1972年美国总统尼克松访问了红色中国，中国大陆上在进行轰轰烈烈的文化大革命，到处是红旗和虚妄的口号。

1965年7月霍金一拿到博士学位就与一个叫简的女生结婚。1970年他得靠四腿的架子才能走路。1972年他开始使用轮椅至今。1985年他访问中国，在北京师范大学做了一次报告后去了长城，爬完长城后回到欧洲就因严重的肺部感染做了气管切开手术，保住了生命，但从此失去声音。此后他依靠为他专门设计的一台语音合成器来说话，通过握在手上的开关控制计算机，一分钟最多可以造一个简单的句子。虽然很艰难，但霍金却十分幽默而乐观地用这一系统进行语言交流，写论文和著作。本书作者张轩中曾经与霍金在2006年有过短暂交流，发现霍金是用眼睛在电脑屏幕上打字，打几个字就会汗流浹背，让人觉得十分可怜可敬。

也许是他写文章不容易，所以读他写的文章，感觉非常干净，一般不给人乱糟糟的感觉。

可能，1998年是霍金最后的创作时期，他写了很多文章，在hep-th查一下那是他文章最多的一年，他在1998年11月的文章是关于ADS / CFT对偶的，他研究了反德西特空间里的克尔黑洞的热性质。2002年霍金大谈膜宇宙，2004年他在都柏林制造新闻，大谈黑洞辐射的信息守恒问题。黑洞辐射的信息守恒问题科普地说大体是这样的：假如进入黑洞的是一个胖子，那么辐射出来的还是不是一个胖子，他会不会变成一个瘦子。

(2)

早在20世纪60年代，英国物理学家霍伊尔在BBC的广播里有档科普节目，每天晚上大讲相对论啊外星人啊，那在英国是一个科普的60年代——类似于最近的中国，有女航天员太空授课，科普受到热捧——那些夜晚是美妙的，震撼人心的，霍金也听到广播，所以1962年快要在牛津大学毕业的霍金，申请去剑桥大学攻读宇宙学博士学位，他心目中的导师就是霍伊尔。但是后来剑桥大学安排给他的导师是一位他从没有听说过的丹尼斯·莎玛（Denis Sciama），霍金将这视作灾难，丹尼斯·莎玛也培养了一批相对论方向的研究生，包括霍金与卡特在内，八卦一句，上海天文台的景益鹏也在丹尼斯·莎玛那里学习过，当时丹尼斯·莎玛已经去了意大利的里雅斯特理论物理研究所，景益鹏做的东西偏天文，也是张轩中在浙江春晖中学毕业的师兄。

1970年的霍金发现黑洞的有趣的动力学性质，如果两个黑洞碰撞并且合并成一个单独的黑洞，围绕新形成黑洞的事件视界的面积比分别围绕原先两个黑洞的事件视界的面积的和更大，这相当于说 $1+2>3$ ，这样的记号不能往死里理解，正如歌德巴赫猜想不是真的要证明 $1+1=2$ 。霍金的这个发现就是面积不减定理，它无疑暗示，在一颗黑洞的事件视界面积和热力学的熵很类似。热力学第二定律说，熵总是随时间而增加。如果把黑洞的面积理解成为熵的话，那么这一切就很漂亮，霍金的“面积定律”，即稳态黑洞的“视界”的面积随时间永远不会缩减，这似乎与热力学第二定律有异曲同工之妙。但黑洞动力学可以当作黑洞热力学吗？当时的霍金还没有这样的意识。

霍金天然地认为黑洞没有温度，那它就不可能有熵——因为物理学家从来没有相信过黑洞是像木炭那样烧红了在天空中，而一直以为它是全黑的，也就是没有温度的。所以他觉得黑洞的视界肯定与热力学的熵没有关系。但是远在美国的年轻人贝肯斯坦（Jacob Bekenstein）有一天

对惠勒教授说：“黑洞视界的面积不只是接近黑洞的熵——实际上就是黑洞的熵。因为.....”

因为什么呢？

假如黑洞存在，就在你的办公室里，你把一杯开水倒进黑洞里，那么杯子里的熵就减少了，这是违背热力学第二定理的，所以只能把黑洞和杯子看成一个整体，熵没有减少，而是跑到黑洞里去了！

惠勒对贝肯斯坦说：“你的想法有点大胆的疯狂，但很有可能是对的，那么你就拿出去发表吧！”于是贝肯斯坦在1973年在《黑洞热力学》一文中正式发表了自己的观点。注意，这个文章的题目看上去是前无古人的，是关于黑洞的“热力学”，不是动力学。这里面有一个在霍金看来很不爽的“热”字。

霍金严重地不相信，他和其他两人立即在1973年2月的《数学物理通讯》上发表了经典的《黑洞力学中的四个定律》的论文，反驳了贝肯斯坦。

这个文章思路很清楚，是霍金那简洁明了风格的写照，也算是广义相对论研究的集大成之作。他完整地写出了黑洞动力学的四个定理。情景完全类似于牛顿的三个运动定律。

黑洞动力学第零定律：稳态黑洞的表面引力在视界上是常数。

黑洞动力学第一定律：稳态轴对称黑洞质量 M ，事件视界面积 A ，表面引力 k ，角动量 J ，角速度 ω 满足一个关系 $dM = (k/8\pi) dA + \omega dJ$ 。

黑洞动力学第二定律：事件视界面积在演化中不会减少。

黑洞动力学第三定律：不可能通过有限次操作把黑洞表面引力降为零。

但是，这四个定律，其实越看越像是热力学定律。

第一定律和热力学第一定律很相似，也就是能量守恒定律，只要把 k 看成温度， A 看成熵就行。第二定律是霍金之前的结果，它不允许单个黑洞分裂成为两个，而且要求两个黑洞碰到一起形成新的视界面积一

定要大于原来面积的和。第三定律并没有严格的数学证明，但是有些很强的证据，它与从旋转黑洞里提取黑洞转动能的彭罗斯过程有关系，彭罗斯过程可以降低表面引力，但是当表面引力越来越低的时候，彭罗斯过程的效率也越来越低，趋于零。这在热力学里就是说，绝对零度是不能达到的，也就是能斯特定理。

但从这样的相似性里还不能断言，这四个黑洞力学公式就是黑洞热力学的定律。因为黑洞是一个绝对的黑体，它温度为0，什么样子的辐射它全能吸收，所以它的熵要是存在，那一定是无穷大。所以霍金他们确定：黑洞动力学和热力学定律的相似只是表面的。

贝肯斯坦发表文章以后，自己也犯嘀咕。他后来回忆说：“在1973年那些日子里，经常有人告诉我走错了路，我只能从惠勒教授那儿得到安慰，他说，‘黑洞热力学是疯狂的，但疯狂到了一定程度之后就会行得通。’”开始霍金根本不把初出茅庐的贝肯斯坦放在眼里，但是，最后贝肯斯坦胜利了！

贝肯斯坦的直觉是正确的，同时他也是幸运的，因为他的想法其实不是最深刻的，甚至于是有一点naive。他要想服众，必须说明一件事情，那就是黑洞不是零温的，这样的话，黑洞因为具有温度，才可能具有有限的熵。这一次还是霍金挽救了他。1974年年初，霍金把量子力学用到黑洞领域，他非常惊讶地发现，黑洞似乎以恒定的速率发射出粒子。

这一次，霍金简直成了神。

以前的经典广义相对论认为黑洞不能发射粒子。但当量子力学加进来的时候，黑洞正如同通常的热体那样产生和发射粒子，这热体的温度和黑洞的表面引力成比例并且和质量成反比，它的辐射谱是热谱，所以辐射不带有任何有意思的信息。但这使得贝肯斯坦关于黑洞具有有限的熵的论点站住了脚，黑洞以某个不为零的温度朝外辐射，黑洞有熵。

(3)

1973年的圣诞节是霍金黄金时代的高潮。他开始用弯曲时空量子场论证了坍缩中的黑洞会有热辐射。这意味着黑洞并不是完全黑色的，这是一个伟大的进展。到了1974年的1月，在他的文章正式发表之前，他把这个想法告诉了莎玛，莎玛把这个消息告诉了彭罗斯，彭罗斯也非常兴奋。马上就有很多别人的文章引用霍金的结果，在当时还没有互联网，但整个研究引力物理的圈子已经轰动了。可见当时霍金对黑洞热辐射的证明颠覆了经典广义相对论给人留下的冰冷印记。这一年霍金才32岁，还算是非常年轻的研究人员。

黑洞居然是热的！

大家全在谈论霍金的工作。英国有一位叫泰勒的理论物理学家在霍金做报告的时候作为会议主席激烈反对霍金关于黑洞热辐射的观点，并且拉着另外一位物理学家当场离开了会场，他认为霍金简直是一派胡言。苏联的泽尔多维奇一开始也持保留意见，他小组里的其他成员自然也不敢支持霍金——因为当时苏联的研究小组，在学术思想上也是非常专制的，学派领袖具有非常高的权威。

霍金的工作在1974年发表在《自然》杂志上。他的证明是用到量子场论的。1976年，意大利的鲁非尼（Ruffini）和他的研究生达莫（Damour）用量子力学重新证明了黑洞具有热辐射。

黑洞辐射的温度和黑洞的质量成反比，所以对于恒星形成的大质量黑洞来说，热辐射的温度比宇宙背景辐射的温度还要低很多，所以没有什么现实意义，但对于人造小黑洞就不一样了，比如两个质子碰撞，如果碰撞的能量足够高的话，碰撞就会产生一个小黑洞，然后小黑洞蒸发，就会产生一些光子，引力子，中微子等——这是人们希望在高能强子对撞机里看到的实验现象，读者们可以等待2015年以后LHC的实验结果。当然人们只能到黑洞的辐射谱，根本看不到碰撞之前原来的那两个

质子的信息，但目前的全息原理认为，原来那两个质子的信息，包含在黑洞的视界之上，只不过不能把这些信息提取出来而已。

黑洞辐射的谱与宇宙微波背景辐射一样，是黑体辐射谱。什么是黑体热辐射呢？在一个夜晚一群人围坐在炉子边上聊天，山河大地，一片苍茫。但大家感觉到身子是热的，这就是热辐射在起作用，黑体谱的意思是发光的炉子本身达到了热平衡——比如太阳的外层，里面的核反应源源不断提供的能量与辐射出去的能量相互平衡，所以太阳辐射就是典型的黑体辐射。黑体辐射的粒子是光子，也可以是其他粒子。

热辐射的光子数目是非常多的，多到要以1mol为量纲，也就是 10^{23} 次方个。这么多的光子当然需要用热力学统计。玻耳兹曼和麦克斯韦等人早已经发展好了热力学统计。对于处于热平衡的系统，可以引进配分函数，然后计算平均能量。在普朗克的计算过程中，因为一个粒子处于能量为E的热力学概率是 $\exp(-E/KT)$ 。但 $E=n+1/2$ （n是自然数，这一步是量子力学的，也是引人入胜之处）。计算配分函数的过程就是高中数学的等比数列求和，而根据配分函数就可以得到普朗克的热谱。

黑洞的热辐射谱也正是普朗克谱。和普朗克一样，霍金也是考虑了量子效应。因为在黑洞的表面，真空的涨落会产生虚粒子对。虚粒子对是一对生死契阔的情人，一般情况下不能分开，要想把这两情人分开，需要外场提供能量，比如电场，或者引力场等。牛郎和织女本来是不会分开，许仙和白素贞本来也不会分开，但一旦外来的黑暗势力介入，情人就分开了。这就是爱情之所以成为千古绝唱的原因。

虚粒子对因为黑洞表面强大的引力场的拉扯而错失对方的手。当其中一个虚粒子被黑洞的强大引力吸引掉进了黑洞，那么他已经与他曾经的情人阴阳两隔，只能是人鬼情未了。

现在假设情人中的女孩已经被黑洞强大的引力场带走，男孩有两种选择。

(1) 跳进黑洞，跟着女孩。

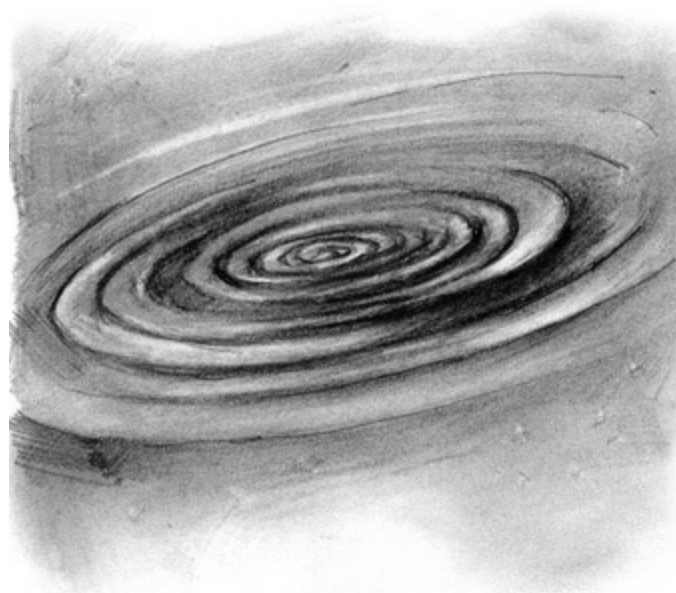
(2) 远离黑洞。

选择(2)方案的男孩跑出来，就成了霍金辐射。这很像电影《大话西游》里的至尊宝孙悟空，他放开了紫霞仙子的手，紫霞仙子进入了黑洞，已经死去，至尊宝孙悟空失去了爱情，就一个人独自走开，也许他确实应该去西天取经，爱情已经没有了，生命还留着，那么干点什么呢？自然是干一些舍生取义普度众生的事情。

霍金辐射就是这样产生的，这个逃离了黑洞的男孩现在没有了爱情的牵绊，相当的自由。有很多这样的男孩子，他们成群结队，翻过一个叫里格-惠勒的山。他们有的正穿过星汉灿烂地奔向我们的地球。

黑洞对他们来说，是埋葬爱情的最美丽的坟茔。相思宛若离草，爱恨渐行渐远。

他们要发光，照亮这个黑洞。



黑洞 绘画：张京

33 宇宙学

(1)

1917年，爱因斯坦想把他的广义相对论应用到整个宇宙，他这个时候是38岁，之前的研究工作几乎全部尘埃落定，他已经到达了他个人研究的高潮时期，并且开始走下坡路了。1917年他还搞了一件事情，那就是激光的原理，要知道这时候连量子力学还没有建立起来呢，而1917年以后，爱因斯坦也不是无所事事，他搞出了玻色-爱因斯坦凝聚，也提出了EPR悖论，所以，爱因斯坦一辈子在这么多领域里做出杰出贡献，爱因斯坦在科学成就上远远超越霍金，论科学成就，如果爱因斯坦得100分，霍金最多只能得40分，所以称霍金是当代活着的爱因斯坦毕竟有不少水分。闲话暂且不表，1917年爱因斯坦当时相信宇宙静态——就是说宇宙是一个永恒的不会变化的一个空间，静态宇宙模型在空间上是一个三球面，所以为了得到静态的宇宙，他必须提供抵抗引力的排斥力，于是他引进了正的宇宙学常数，来产生排斥力。爱因斯坦的静态宇宙是经受不起微扰的，也就是说，一个胖子在地球上跺脚，可能引起宇宙剧变，天崩地裂。所以，这一次爱因斯坦在物理上表现得非常粗糙，简直显得有点业余——他的宇宙模型就好像是放在马鞍面上的篮球，随时都会滚下来。不过2005年在纪念爱因斯坦提出相对论一百年的时候，大物理学家温伯格写了一篇广为流传的文章，题目叫《爱因斯坦的错误》，在文章中，温伯格认为爱因斯坦提出静态宇宙模型在当时是非常正常的。

1917年爱因斯坦提出静态宇宙模型，这是他把广义相对论运用到宇宙学的第一个实际运用。可是没几年，天文学家Slipher发现了远处有不少星系似乎都在离开我们银河系。这个观测似乎表明，宇宙空间正在膨

胀。

爱因斯坦这个时候开始意识到他的静态的死气沉沉的宇宙模型应当放弃。尽管这样，爱因斯坦作为宇宙学的创始人他的工作为后来的宇宙学理论的发展奠定了立法基础：首先是宇宙空间演化的动力学服从广义相对论引力方程的规律，也就是说，宇宙这个大时空其基本结构由物质场决定，其次是宇宙空间存在所谓宇宙学原理，也就是说，此地与彼地没有区别。当时这个宇宙学基本法的两点都并没有事实基础，而只是猜测性的假设。

1922年，俄国物理学家弗里得曼在爱因斯坦宇宙学原理的假设下建立了膨胀的动力学方程。因为当时在天文上还没有任何观测能证明宇宙的空间部分正在膨胀，所以这个理论被束之高阁。一直到了七年之后，1929年，美国天文学家哈勃发现：其他星系在向远离我们银河系的方向运动，其离开的速度（通过光的红移来计算）与距离成正比。

（2）

今天被称为大爆炸理论的一系列论文是伽莫夫等人在1949年前后两三年里发表的。

杰出的天文观测家哈勃也犯了个不小的错误，在他发现的膨胀规律中把速度与距离之间的比例系数（即哈勃常数 H_0 ）的值是 $H_0 = 500 \text{ km/s} / \text{Mpc}$ 。按动力学理论一推算，这意味着今天的宇宙年龄约是20亿年。当时已经知道，地球的年龄是45亿年，地球诞生在宇宙之前，这当然是不能接受的结果。此后20年宇宙学几乎没有进展，一直到了1949年。

伽莫夫在1949年问了一个问题：一切星系形成前的宇宙是什么样的？由于星系形成过程的不可逆，要用物理规律来追溯回去是不可能的。伽莫夫猜想，那时的宇宙应是一大片均匀的高温气体，根据自引力不稳定性理论，均匀气体中的小密度起伏是会发展出物质结块的。至于

事实是否如此，事先无法知道。

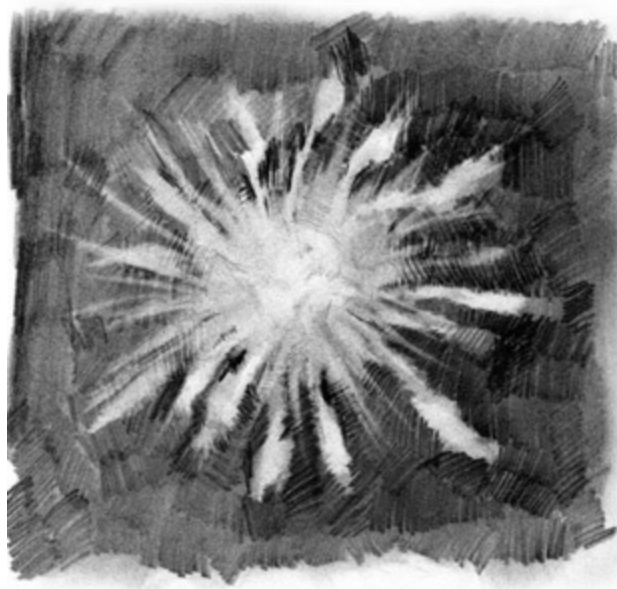
他首先研究了化学元素的起源问题，按伽莫夫的理论，宇宙年龄为1秒的时候，宇宙中粒子的热动能约为1MeV。那时热碰撞足以使原子核解离，所以化学元素应当是此后产生的。当时已经初步知道，氦在今天宇宙中的丰度约高达1/4，它是除氢之外丰度最高的元素。虽然恒星中心的核燃烧会产生氦，但估算表明远不可能产生这么多。伽莫夫发现宇宙早期的元素合成确实足以产生丰度达到1/4的氦核。

伽莫夫研究的另一课题是微波背景辐射。在原初核合成后，宇宙气体处于等离子状态，随着宇宙膨胀，气体将绝热降温。在温度降到1万度以后，最后粒子的平均热动能只剩1eV，这个时候原子核与电子将开始结合成中性原子。一旦电中性的原子成了主要组分，原来处于热平衡等离子态下的光子组分就失去了碰撞的机会。这就是光子的退耦，退耦后的光子当然会永远存在下去，它就是现在人们讲的宇宙微波背景辐射。背景光子虽然失去了碰撞，但它过去的热平衡分布（Planck分布）将保持，其等效温度将继续随宇宙膨胀而降低，因为宇宙空间的膨胀会把所有波长相应拉长，根据维恩位移定理可知等效温度相应降低。伽莫夫当时估计出宇宙微波背景辐射的温度在10K左右，其波长主要部分在微波波段。现在回顾去看，当年伽莫夫的理论（后来被叫做大爆炸理论）做得非常精美。可是当时的学者却很不以为然，因为在他们看来，连宇宙膨胀都不可信，用它推至这么早的宇宙还能信吗？

（3）

20世纪60年代初，普林斯顿大学的迪克和皮伯斯重新研究了这方面的理论问题，迪克是标量张量理论的提出者——那是一个与爱因斯坦引力理论有竞争的理论，他是一个关心相对论的实验检验的人物，皮伯斯2006年访问过清华大学，他现在宇宙学上的地位很高，他以前是迪克的

学生。当时他们两人打算制造一个微波探测仪器，来探测宇宙微波背景。但是，他们正在忙碌的同时，意外发生了。



大爆炸 绘画：张京

1965年离普林斯顿不远的贝尔实验室的工程师彭齐亚斯和威尔逊意外地发现了宇宙最早的光。他们在波长为7.35厘米的长波段发现了温度为3.5K的不明信号——这个温度是根据电子学里的纳奎斯特定理估计出来的。这个信号非常特别，就是无论你怎么改进探测仪器，它永远如影随形，不可消除。这个信号甚至与时间无关，与空间无关。也就是说，在任何季节，这个信号都存在，在天空的任何方向，这个信号也存在。彭齐亚斯和威尔逊完全不懂宇宙学，他们刚开始以为，这事情真是见鬼了，真是越怕鬼越见鬼。但他们还是把他们的观测结果写了一篇1000字的文章发表出去了，意思在排除了微波天线上的鸽子粪以后，这些信号依然存在，特别有寻根究底精神的彭齐亚斯和威尔逊排除了来自天线本身和地球近处的可能，指出它是来自远处的辐射背景。这需要得到大家的解释。论文已经足够短了，但没有想到的是，这个论文为他们赢得了诺贝尔奖。这真是机缘巧合，冥冥中自有天注定。迪克和皮伯斯也在同

一期的《天体物理杂志》上详尽地讨论了彭齐亚斯和威尔逊发现的信号的宇宙学意义：这说明大爆炸宇宙学不是伪科学，这事情太大，人们对此持很谨慎的态度。关键点是怎么更可靠地证明所发现的背景辐射和大爆炸预言的背景辐射是一回事。因为要编个理论来解释已肯定存在的背景辐射并不难，大爆炸理论的坚决反对者多年后还在做这样的事。为澄清这疑问，人们首先想到可以做的是：从各个不同波长（频率）作同样的探索观测，希望既能找到它，又能由实测强度推断出同一温度在3K左右。这一步进展很顺利，当在地球上在10来个波长上都测到了接近相同温度的背景辐射后，瑞典皇家科学院有了信心，把1978年的诺贝尔物理学奖授予了彭齐亚斯和威尔逊（当时伽莫夫已过世，不能得奖了）。

当然，为了更加严格地验证背景辐射确实是黑体辐射谱。1989年，美国宇航局（NASA）曾发射过一颗宇宙背景探测者卫星（COBE），证实了这个结论，因为地球上的大气对电磁波有吸收，高精度的探测必须在大气层外，所以发射卫星是最好的选择。其主要目的之一就是做这件事。COBE在0.1~10mm之间的三十几个不同波长上安置了微波接收器，对背景辐射做精确测量，卫星上天不久，辐射谱的观测结果就得到了，用黑体辐射公式来拟合，竟可定出四位有效数字的辐射温度，宇宙微波背景的温度全是2.735K，这也充分表明其辐射源的热平衡程度很高。至此早期宇宙是一个高度热平衡的均匀气体已无可置疑了。1992年COBE还观测到了宇宙微波背景辐射在不同方向上存在着微弱的温度涨落。这个结果被霍金认为是人类科学历史上最杰出的发现之一，因为只有均匀的宇宙背景里找到涨落，我们的星系和生命才可能形成。

宇宙微波背景辐射是大爆炸遗留下来的目前唯一可以观测的遗产。

对历史学家来说，考古是在发掘遗产，因此现在有很多历史学家想挖掘秦始皇陵墓。但现在中国政府还是禁止实施这个计划。

对物理学家来说，宇宙背景辐射也是这样一份遗产。

第二部分附录

1. 论用初中数学的方法实现相对论在中国的普及

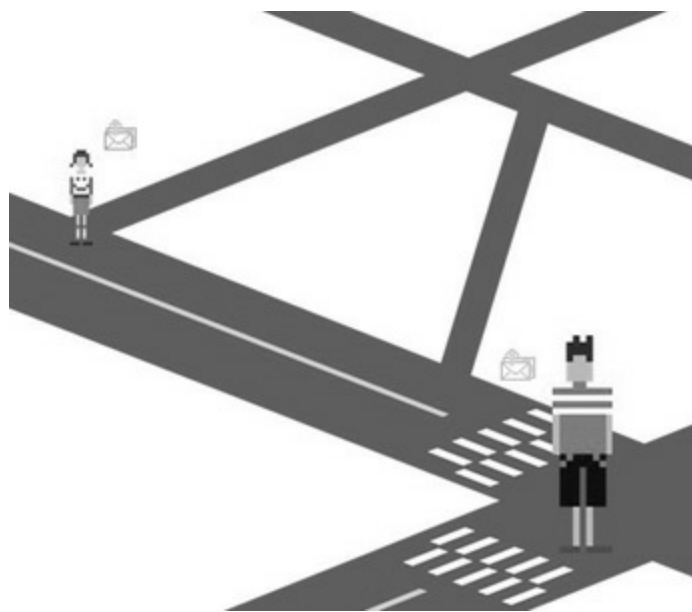
在2008年广西师范大学出版社出版了我的科普书《相对论通俗演义》以后，很多读者对我说过：“你写的书一点也不通俗，我读了以后还是不懂相对论到底是什么！”这让我非常沮丧，也不得不承认自己的书确实有很多不通俗的部分。所以在2008年以后，我开始摸索如何在中国实现相对论的普及，并且与其他几个年轻人一起组织了一个爱因斯坦学校，专门来普及相对论。这个爱因斯坦学校曾经采访过著名的相对论专家赵峥教授、梁灿彬教授，也采访过翻译过《爱因斯坦文集》的范岱年与许良英先生。在爱因斯坦学校秉承的“培养中国的爱因斯坦”的宗旨下，我个人还不断地深入到大中小学校进行科普讲座，普及相对论。听过我的科普演讲的对象包括在北京师范大学给教育学部的吴岩教授所带的文科研究生，果壳网中的文艺青年，江苏的南菁中学的高一学生，浙江的竺可桢中学的初中生，以及北京的水晶石教育的艺术家等——这些人几乎全没有高等数学的基础，但同样可以理解到狭义相对论的思想精神。在这些单位我经过多次的讲演后发现，要在中国普及相对论的基本思想，让相对论像进化论一样做到妇孺皆知，其实只需要初中的数学物理水平，就完全可以弄懂。

下面我就来简单介绍一下用初中数学的方法，如何来讲解狭义相对论。

其实，相对论主要的研究对象是时间与空间。时间与空间组合起来构成一个客观的绝对的物理对象——时空。原则上时空不能用一张二维的“图”来表示，但是如果被研究的物体运动总在同一条直线上，则我们可以选用一个二维的坐标平面去标记它的运动。这时候纵轴标记时间 t ，

称为时间轴，横轴则是物体运动的位移。这就是时空图。

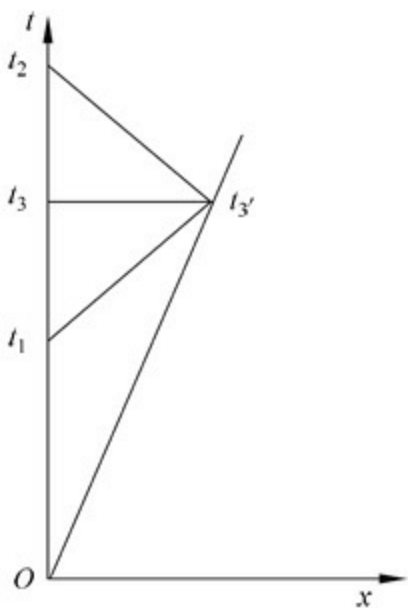
先想象一个每天都会发生的小故事。一个戴着手表的男生和另外一个戴着手表的女生在街头吵架了。接着女生让男生走，于是男生就沿着大街匀速离开了（为了考察到狭义相对论的效应，我们假设男生的运动速度很快）。过了一会儿，女生忽然后悔了，于是赶紧掏出手机给男生发了个短信：“亲爱的，对不起，回来吧。”于是男生收到短信后，原谅了女生，马上给女生回了一个短信：“好的，我马上回来。”



街头时空图

这个故事在时空图上是这样的，我们让坐标原点落在男生和女生吵架的街头，假设此刻他们的手表是刚刚对过时的，一开始都调整为零。然后男生沿着大街以速度 v 离开，于是我们可以把大街设为 x 轴。在时空图上，男生的轨迹是一条直线。而女生虽然站在街头没有移动，但是时间却是在流逝的，所以女生在时空图里的轨迹是和时间轴重合的直线。——我们把这两条直线分别叫做男生和女生的世界线（world line），它表示男生或者女生经历的所有事件的一个集合，世界线的长度就是男生与女生所经历的真实的时间——记住，在爱因斯坦的相对论中，时间是

私有的，与脉搏一样，每个人的时间都走的不一样快。当女生的手表显示 t_1 时刻，女生开始后悔，于是女生发了道歉的短信——短信以光速传播，在男生收到短信的时刻，男生的手表显示的时间是 t'_3 ，女生手表显示的时间是 t_3 。如果男生收到短信的刹那就给女生回复了短信，那么当短信以光速传到女生的手机，女生的手表显示的时间是 t_2 。狭义相对论的一个重要结果，就是发现 t'_3 与 t_3 是不一样的。这个过程用下面的时空图来表示。



女生为参考系的坐标图

图中 Ot'_3 是男生的世界线， $t_1t'_3$ 和 t'_3t_2 是短信的世界线

以前已经说到，根据狭义相对论，每个人都有自己的时间，称为“固有时”。在上面的例子中，时间轴上的时间实际上是女生的固有时；而男生手表显示的时间才是男生的固有时，它等于男生的世界线的长度。因此，如果我们想要知道 t'_3 和 t_3 差了多少，我们就需要知道他们之间的关系。

在上面这幅图中，是两个在时空中的“相似三角形”，或者更确切地说，是物理上的“相似三角形”（在这里作者必须强调，这个相似不是初

中数学中学过的欧几里德几何意义上的相似，而是在时空的几何学上的相似。作者曾在中国科学院物理所给一些物理学家讲过这个，大家一开始也都误会成这个相似三角形是我们通常的相似三角形了，其实，这是物理相似三角形，用到短消息是光速运动的，而回短消息的那个男生是在运动的，但他回出去的短消息还是以光速运动，这背后其实隐藏了狭义相对论的光速不变原理。而女孩子发短消息与男孩子发短消息这2件事情在物理上是等价的，这说明物理学与性别无关，物理学不是社会学，这物理相似三角形无非说明，以女孩子为惯性参考系来描述物理规律，与以男孩子为惯性参考系来描述物理规律，这在本质上是没有任何区别的，这叫做狭义相对论原理。）。之所以说是“物理上”的，因为你也看得出来，它和我们一般在欧几里德几何学里所学到的相似三角形不太一样。但是我们怎么知道它们是相似的呢？这就要靠物理上的判定了：女生给男生发短信和男生给女生发短信在物理上是完全等价的。也就是说这两个操作在时空图上具有等价性。也许你可以自己画一个在以男生为参考系的坐标系，这时候在男生看来女生才是相对于他远去的——这样你就可以发现，的确是两个相似三角形。那么由相似比，我们可以得到 $\frac{t_2}{t'_3} = \frac{t'_3}{t_1}$ ，即 $t_1 t_2 = t'^2_3$ 。注意到男生与女生在 t_3 时刻的空间距离 L ，它既等于男生离开的速度 v 乘以女生手表流逝的时间 t_3 ，也等于短信的传播速度光速 c 乘以短信以光速传播需要的时间。因而我们有

$$L = vt_3$$

$$L = c(t_3 - t_1)$$

$$L = c(t_2 - t_3)$$

消去 t_3 ，并且整理我们可以得到 $t_1 t_2 = L^2 \left[\left(\frac{1}{v} \right)^2 - \left(\frac{1}{c} \right)^2 \right]$ 。利用 $L =$

vt_3 ，还可以得到 $t_1 t_2 = t^2_3 \left[1 - \left(\frac{v}{c} \right)^2 \right]$ 。再利用前面提到的物理相似三角

形的相似比 $t_1 t_2 = t'^2_3$ ，就可以得到 t'_3 与 t_3 的数学关系了。

从结果我们可以看出，时间“膨胀”了，这正是相对论的一个重要结果，也是著名的洛伦兹变换中时间的洛伦兹变换。

因此，我们可以看到，使用初中数学中所学到的相似三角形知识，以及初中物理中最简单的路程等于速度乘以时间的物理知识，我们就可以很快地得到不同运动速度的物体之间的时间走的快慢。这个推导过程笔者曾经在多次科普讲座中给观众演示过，几乎大部分观众都能够听明白，从此就了解到狭义相对论的思想精髓。

相对论诞生于1905年，至今已经有100多年的发展历史，但普及相对论的工作却一直步履维艰，早在诗人徐志摩还在世的年代，他的大舅子张君勱曾经给他一本相对论的书叫他看，他看了以后囫圇吞枣，最后也写过一篇科普相对论的文章《安斯坦相对主义——物理界大革命》发表在《改造》杂志上。但可惜的是，徐志摩本身没有搞懂相对论，所以他的科普文章写得一塌糊涂，科普效果很差。一百年后，通过笔者最近所做的科普工作，也许普罗大众终于可以真正搞懂狭义相对论究竟是什么了。

2. 中国式相对论

记得有一位不知名的人曾经说过，相对论好像是一个金字招牌，这话很多人不爱听。据说曾经有一个数学家好像是拉普拉斯还是拉格朗日，和一个哲学家在吵架，吵架的时候这个数学家说了如下类似的话：因为 $x^3 + y^2 = 1$ 是一个椭圆曲线，所以可以证明你是一个傻×。哲学家马上被吓得哑口无言，面色惨白，败下阵来。

相对论当然也有同样的功效。

但相对论也具有数学的结构，比如恩斯特方程就是一种把稳态的轴对称的爱因斯坦方程求解出来的数学方法。上帝保佑吃饱了饭的人们，

当吃饱了饭，一边喝咖啡一边看看这些数学结构，人们一定会很陶醉。因为这些结构与人无关，有点God given的意味。

我在上大学的时候，那是在2002年吧，也许是2003年，去听梁灿彬老师讲微分几何入门和广义相对论。当时上课用的书还是梁老师的第一版书，就是北京师范大学出版社出版的。书是在物理系买的，因为梁老师害怕我们这些穷学生为了学相对论破费了，所以他委托物理系的人低价卖给我们。有时候他甚至愿意为一个同学买书亲自去跑一趟出版社，这种情操一般人肯定是学不来的。梁先生有时候帮学生买书，好像是骑着他的自行车去的，因为他现在已经是头发全白了，当年也已经是一位老人，长辈对晚辈有时候就是这样的，我们都是孩子而已。

其实梁老师在物理系讲这个相对论课程是有点历史了的，所以在2001年，也许我就听过一次他的讲课，那节课我当然是听不懂，那时候我才上大学一年级，猥琐地在教室后面听着，坐在前排的都是一些研究生。我记了一些笔记，后来翻笔记的时候，才知道当时讲的是洛伦兹群和托马斯进动这一类的东西，用的是李群的语言。当时我完全不懂相对论，所以那一晚上相当于坐了一次飞机。

2003年左右正式成为梁老师课堂上注册的学生，我决定从头开始学起。当时一起去上课的人是很多的，同年级的同学不少都要去听一听，听上那么一耳朵，有的人就跑了。

第一节课自然是讲流形，流形上的坐标是流形上的一个点 p 映射到 R^n 上产生的，这里射来那里射去，这完全是数学语言，物理系出身的学生多数一开始还真不习惯。当时大家就全懵了，于是，上了几次课，听课的人数也被映射到了一个不动点——大约10个人，从50个到10个，大约需要半年的时间吧，但这个课程预计是要连着上两年的，所以10个剩余的人是真正喜欢相对论的。

梁老师非常关心这些学生的作业情况，所以每周有答疑课，由他亲自在物理楼一间办公室里回答学生的问題。他还非常关心学生的学习进

展，有一次，他打电话到我们宿舍，我接了电话以后，他说他要找周地宝同学，问他是不是住在隔壁宿舍，好像是讨论作业的问题，还是讨论为什么上次课没有去听，反正我也有点忘记了，毕竟是10年以前的事情了。我因为在一次答疑的时候，问了他一个关于环面能不能平坦地嵌入到3维欧氏空间里的问题，他一下子没有回答上来。有一次我正好在校园里走，梁老师骑自行车去邮局，在路上遇见我，下车来叫上我，告诉我上次我问的那个问题的答案。

因为相对论这个课程实在是太冗长了，我有时候因为忙别的事情或者听不太懂，偶尔也要逃课，但毕竟勉强还是把2年给坚持了下来。而我的很多同学，现在有的搞凝聚态物理，有的搞天文，有的搞实验物理，他们也许永远失去了搞一搞相对论的机会，尤其是搞一搞广义相对论的机会。很多人以曾经搞过物理为荣，只有一小部分人能以曾经搞过广义相对论为荣。这就是生活，生活总是这样，若湍流一样激流奔涌，年轻的时候往往不知道学什么才真正有用，被生活裹着走了。

梁老师的课教给我们很多东西，每个学生都从他那里学到很多知识。我当然是不清楚别人学到了什么，对我自己来说，每当我在企业里看到一些管理人员在搞一个高斯分布，然后说六西格马管理如何有效，当我看到那些工程师在盲人摸象地做一些物理实验，我就知道，我和他们不同，因为我自己从梁老师那里学到了一点微分几何的皮毛，从而似乎让我有勇气做我自己了。

“你觉得你最擅长的是什么？”一个同事问我。

“微分几何吧，也许是相对论。”我说。

“微分几何？相对论？”他说。

尤其是想到二维的曲面上外尔张量处处退化，它是局部共形平坦的，我总能感觉到一种迷醉。附庸风雅的7年，就在这样的迷醉里度过了。

听完梁老师的课，这时候已经是大四了。大四在大学里是一个很尴

尬的年纪，因为大四的时候，你会觉得自己真的什么也没有学会，并且你会觉得自己已经不再年轻了。在这个时候，物理系里贴出了一张大大的喜报，说北京师范大学物理系毕业的校友桂元星教授，因为在引力方面的工作，得了一个什么奖。好像是找到了爱因斯坦场方程的一个度规？反正我也忘记了，但当时看到那张红纸，看到了中国人还真能找到爱因斯坦方程的解我觉得这就是厉害的。

有的人晃悠悠过一辈子，有的人则生下来就有强烈的目标感。大四的时候，同学们自然是自个儿打算了。找工作，出国，考研究生，无非是3条出路，创业似乎是不太可能的——至少对很多人来说，创业需要的第一桶金是很难找到的。

我也在混沌之中，因为我也算是学过一点相对论，于是考了物理系马永革老师的研究生。结果就考上了，于是就接着读和相对论有关的领域，当然顺便也学点圈量子引力。大四和研究生一年级的時候，赵峥老师讲黑洞和物理学历史，我自然也去听听的。黑洞热力学也算是听了一些，最近记得的已经很少，只记得说黑洞有一个负的热容，并且，黑洞的熵和面积成正比，这玩意儿自然是大家都晓得的。最近因为工作的关系，像一个工程师一样干活，研究了一下液体球的表面张力的能量，发现这个表面能似乎也与球的面积成正比。所以，当在工作的时候，我和同事们假装在谈论小液体球，其实满脑子是觉得自己在谈黑洞熵。

当你把一个黑洞想象成一个液体球那样会晃动的玩意儿，你也许会遇见黑洞微扰，钱德拉塞卡因为搞黑洞微扰发现了一些美妙的结构，所以他说：“广义相对论是真理。”

赵峥老师的物理学史课，是一门大课，很多外系的人也听，所以每次讲课是在北师大那个钟敬文讲堂里，那里面至少可以容纳400人吧，几乎是场场爆满，他的物理学历史讲得好，简直比易中天讲的还要生动一些。因为易老师讲的是人性和历史，而赵老师讲的物理学史，不但有人性和历史，甚至还有物理。反正我是听来了一些东西，然后鹦鹉学舌

地写进了《相对论通俗演义》里，比如说开普勒死在讨债途中，第谷有半个金鼻子，很多很多，都是听赵老师讲的。听赵老师的物理学历史课，很难讲不是一门艺术课，每次听课都好像是在看一场电影。

西北望，射天狼。

年轻的时候，能听到良师讲课，这是幸运的。

赵老师不但给学生们讲课，有时候还请学生吃饭，甚至有时候要给学生发点钱，这样的日子，自然是好。那些年，似乎是人与花皆好，而且赵老师有时候喜欢和学生们谈论历史和现实，我们总是听到很多以前的故事。

刘辽教授因为年纪大了，所以很少参加引力组的活动，他也不认识很多年轻学生。但我还是记得有一次，大约是大四的时候吧，刘辽教授家里的电脑坏了，需要找一个学生去修一修，裴寿镛老师叫我去修，我就叫了同学裴彤去了刘辽教授的家里。

电脑好像是网络连不上，反正也忘记了，只记得刘教授家中有一些发黄的藏书，一张大床，墙壁上贴着几句诗。这几句诗似乎非常辛酸悲怆的，我也忘记了，反正纸张很黄，而且字迹有点沧桑感，这就是历史吧。

我好像还和刘教授说了一些关于李群的书的莫名其妙的对话。也已经忘记掉了。

最近参加刘辽教授80华诞的聚会，得到一本《刘辽文集》，里面不但有学术文章，而且还有几首诗，我倒是看了几首诗歌，其中一首自然是很妙，写的是年迈以后路过巴东想起年轻时候遇见一位少女的经历。这诗歌还被凌意教授转载到他的博客上，看来这个诗歌确实是有生命力和自我繁殖的能力。

读了研究生以后，也听周彬老师讲李群，不过只有一个学期的课程，也没有学懂，感觉很是遗憾。后来研究生毕业，倒是周老师辗转介绍到一个工作机会给我，我才有了一次打工的机会。这人生机缘巧合真

是很难预料。中国式的相对论其实是由很多很多人物的贡献组成的。前几天我读王永久老师的巨著《引力论和宇宙论》，感觉此书甚厚。我看吴忠超教授翻译的霍金的著作，也见过吴老师好几次，虽然他不会认识我，但我从小读的湖南科技大学出版的《时空本性》，知道扭量一词，都是吴老师翻译的吧。而其他更多的人，就不写了，因为中国式相对论是一个很大的题目。没有人有资格拿这样大的题目写文章，我写下一点自己的回忆，是为了怕自己忘记。记得有一个人曾经说过，搞相对论的人多数是投机主义者。如果说投机是为个人的私利，相对论也许是不值得投机的，因为成本很高，如果投机是为这个社会还有理想，那中国的相对论领域确实出现过一些众所周知的革命者。

还是以《大话西游》的片尾歌词做结尾吧，是才华卓绝的香港音乐人卢冠廷的《一生所爱》，歌词第一句就是“从前现在过去了再不来……”，曲调中透着莫名的惆怅和忧伤。

参考文献

- [1] 赵峥. 探求上帝的秘密 [M]. 北京: 北京师范大学出版社, 2009.
- [2] 张轩中. 相对论通俗演义 [M]. 桂林: 广西师范大学出版社, 2008.
- [3] 裴寿镛. 量子力学 [M]. 北京: 高等教育出版社, 2009.
- [4] 赵峥. 黑洞的热性质和时空奇异性 [M]. 北京: 北京师范大学出版社, 1999.
- [5] 梁灿彬, 周彬. 微分几何入门与广义相对论 [M]. 北京: 高等教育出版社, 2006.
- [6] 刘辽, 赵峥. 广义相对论 [M]. 北京: 高等教育出版社, 2010.
- [7] 侯伯元, 侯伯宇. 物理学家用微分几何 [M]. 北京: 科学出版社, 2007.
- [8] 李淼. 超弦史话 [M]. 北京: 北京大学出版社, 2005.
- [9] 徐一鸿. 爱因斯坦的宇宙——老人的玩具 [M]. 张礼, 译. 北京: 清华大学出版社, 2006.
- [10] 关洪. 原子论的历史和现状——对物质微观构造认识的发展 [M]. 北京: 北京大学出版社, 2007.
- [11] 柯善哲, 等. 量子力学朝花夕拾 [M]. 北京: 科学出版社, 2007.
- [12] 霍金. 果壳里的60年 [M]. 李泳, 译. 长沙: 湖南科技出版社, 2007.
- [13] 霍金, 彭罗斯. 时空本性 [M]. 杜欣欣, 吴忠超, 译. 长沙: 湖南科技出版社, 2006.

- [14] 徐仁新. 天体物理导论 [M]. 北京: 北京大学出版社, 2006.
- [15] 曾谨言. 量子力学 卷一 [M]. 北京: 科学出版社, 2007.
- [16] 曾谨言. 量子力学 卷二 [M]. 北京: 科学出版社, 2007.
- [17] 徐一鸿. 可畏的对称 [M]. 张礼, 译. 北京: 清华大学出版社, 2005.
- [18] 吴振奎. 数学解题中的物理方法 [M]. 哈尔滨: 哈尔滨工业大学出版社, 2011.
- [19] 刘慈欣. 三体 [M]. 重庆: 重庆出版集团, 2008.
- [20] 赵凯华. 定性与半定量物理学 [M]. 北京: 北京大学出版社, 2008.

跋

(1)

“量子力学之父”玻尔年轻的时候，那大约是在1912年，也就是辛亥革命的时代，中国还没有人研究量子力学，玻尔解决了氢原子的能级问题，得出氢原子的能级问题与正整数 n 的平方的倒数成正比。玻尔的思路是非常自然的，不会让任何人觉得吃惊。这个思路的核心就是所谓“对应原理”——其实就是一个他自己发明的原理，目的是为了解释自然现象，这个原理有一些朦胧的思想，大概就是说电子作圆周运动的频率在经典极限下会与电子所发出的光的频率一致，这里面充满了共振的思想，对读物理的人来说，可以说也是一种模糊的直觉。这个对应原理成为后来的海森堡最厉害的思想武器。实际上，对后来者来说，对应原理是一种真正的物理方法，换句话说——这是物理学家做事情的一般方法，靠直觉去做猜测，然后由实验进行检验。

在玻尔的原子模型里，电子还是按照卢瑟福的模型在不同的圆形轨道上运动，但这些轨道可以用自然数 n 来标记。其实轨道是不存在的——因为人类也观测不到那些轨道。然而，玻尔的思路是非常完整的，在经典力学里就可以知道，不同轨道的能量不一样，可以把第 n 个轨道的能量记为 $E(n)$ 。因为 n 是一个整数，所以 $E(n)$ 是一个未知的数论函数——所谓数论函数，就是自变量是整数的那些函数。玻尔认为，电子在每一个轨道上是稳定的，但可以在不同的轨道之间相互跳跃，这就是量子力学的基本思想精髓。这个跳跃过程被称为跃迁——类似于股票市场中的每天开盘的时候指数的那种跳跃，比如，今天的上证指数到了收盘的时候已经有了一条轨道，收盘在2490点，那么，明天早上开盘不一定是在2490点，有可能跳空低开，比如在2440点开盘。那么跳空低开

所引起的市值蒸发，就相当于原子从高轨道能级向低轨道能级跃迁时候所放出的光的能量。

因此，一个恰当的比喻就是，股票市场中的股票总市值总是存在不确定性的，这个不确定性就好像是一个量子力学系统的能量的不确定性。

当然，还有其他的问题值得思考，那就是一个处于激发状态的原子，它到底能维持多长的时间一直停留在激发状态。如果我们继续那个股票市场的比喻，从中国的A股市场来说，如果我们认为在2007年上证指数能待在6000点以上的状态被称为激发状态的话，那么这个停留时间不会太长，这个时间对应的就是一个原子系统的激发态的寿命。

一般来说，对一个原子的激发态的寿命的估计取决于它向下面的能级跃迁的概率。这个概率是可以通过量子力学计算出来的，概率越大，则它在激发状态所呆的时间也就越短。同样道理，在股票市场中，一个高市值的市场必然具有很大的向下走的概率，所以上证指数能停留在6000点以上的时间总是很短的，如昙花一现，可惜的是，至今还没有一个完整的金融学的理论可以计算出这个概率。

而我们早可以计算出，一个激发态原子的寿命，大约只有 10^{-8} 秒。

(2)

而我写这本书的时候，精神断断续续一直处于激发状态的时间，一直持续了5年。

当辛亥革命一百周年的时候，我大致写完这本关于量子力学和广义相对论的书的草稿。写完这本书的那一年，我也刚好30周岁了。人们都说，三十而立，我于是想回顾一下我是如何走过自己的青年时代的。

在我的故乡绍兴，有一些年轻人受到鲁迅的影响，鲁迅是辛亥革命时代出现的一个文学巨子，他看破人世的悲哀，伸出瘦弱的胳膊抵抗旧

中国车轮的沉重惯性，他看到麻木的人性与虚妄的革命，手里只有一支笔的鲁迅总感觉到他所处的时代没有希望，人民也没有希望。对我个人而言，在辛亥革命的前后，鲁迅的笔下有三个人，栩栩如生。闰土出生在农民家庭，曾经年少无忧虑，长大后一无所有，知道稼穡艰辛，却无力去探究稼穡艰辛的根源到底是什么。阿Q是一个无家可归的人，他渴望革命，却想不清楚到底什么是革命，他所看到的革命队伍也是鱼龙混杂的，其实也是一场闹剧。那个在酒楼上的吕纬甫，这个知识分子那几年在人世飘零如一叶孤舟，最后变得模模糊糊，敷衍敷衍，就这样过日子。

在我的情怀里，在我成长的道路上，我看到鲁迅，也看到这些人物，我与他们是同命运的人，千万的中国年轻学子也都与他们是同命运的人。

我并不是要在这里单纯缅怀鲁迅，也不是单纯想说社会满面疮痍，而只是想说，当我写这本书的时候，我将这本书当作半本小说在写，我自己也正在虚构那些物理学英雄历史上的人物形象。

小说曾经是我的最爱，我在春晖中学读高中的时候，就曾经写小说，发表在学校的《春晖报》上，刚进了北京师范大学读书，就在物理系刊《求索》上发表了一些小说，现在看来，当然幼稚可笑。但我这次在本书的创作过程中，还是使用了一点小说的伎俩，虽然全书总体上文体似乎不太统一，但读者们一定要注意，此书是有小说色彩的，并不是完全的物理学史，大家就不要用历史学家的精确的眼光来审视，然后想着来批判我一番，那没有意义——我只不过是想在中国把科普与小说结合起来，梦想写几本类似《三国演义》的书而已，再说金庸也不是历史学家，但他写的历史背景的小说有那么多人读，那就好了——我的理想也是一样的。

在写这本书的草稿时，我花了差不多5年的时候，断断续续，一直以来，我一直鼓励自己要做重要的学问，去思考比较重要的技术性问题，比如BSD猜想费马大定理，比如Connes是怎么证明Morley三角形的，比如Mathieu的离散群和孔采维奇的月光魔群。因此去写一本轻飘飘的科普书，我总觉得会缺少技术性的内容，于是我一直在我的写作中保持一种对深邃性的自我追求，我希望我的读者是如我一样，能读懂深邃的数学物理。但考虑到科普书本身的销售问题，最后，我不得不把这本科普书写成半科普半文艺的物理小说的形式，这在某种意义上是我个人的一种探索，因为我想走别人不曾走过的道路。

赵峥教授在本书的序言里提到，“翻开历史的长卷，我们看到‘自古英雄出少年’，青年人应该有志气，有抱负！”

我现在虽然已经如《人民日报》最近所描述的大部分中国青年那样已经充满了“暮气”，但我也是确实曾经年轻过的。

我年轻的时候，对相对论有浓厚的兴趣。我随便从电脑里翻出一点以前的手稿（大概记录了2003年我读大学三年级时候的生活），就找到了自己为什么现在无知到了只懂相对论的原因。

在2003年5月7日的时候，我写了一点手稿，算是对自己的记录。

Sars在北京城里早就有了，我坐831的公交车去ww她家里家教，路上看见很多戴白口罩的人。等后来北京城里的人都确定了这件事，我们就停课了。

停课的几天我每天上网，感到极其痛苦，因为日复一日。以下是一些琐忆：

1. 先复习了一下schwarzschild解的一些性质。

发现时空的2部分是不连通的，它们分别用schwarzschild解描述。所以发现没有真实的时间。黑洞内部是非静态的.....

2. 找到了kluza——klein理论的资料。

但这个电子文档是ps格式的，打印店全关门了，在没有办法的时候发现还有一家的门虚掩着。进去，却发现人家只能打开pdf文件，因此不能打印。于是回来研究如何把ps格式转化为pdf格式。

3. 再次阅读韩寒的书《像少年啦飞驰》。一年前在火车上消磨时间，重新看了半本，觉得只有那些傻子才觉得这文章不如《三重门》。老枪他们给盗版公司写书.....不断赚钱，不断花钱，过的循环的生活，就是生活的意义所在。

4. 每天晚上看4集《寻秦记》，看到赵倩的死正好一集完，音乐很感人，我都要流泪了。等看完全部《寻秦记》，觉得“人可以回到过去，但历史无法改变”，也就是说存在闭合类时曲线，但不能破坏物理上的因果性。也就是说弑母悖论不成立，因为这个悖论可以用时序保护猜想去解释。

5. 有些问题是不能问的，比如上帝是万能的，那么上帝能不能创造出一块它举不起来的石头。

这个问题的答案我觉得应该是：上帝能够创造那样的石头，但上帝不是万能的。

6. 写了一本书的第5章。发现自己应该考虑一下如何在年底写一个跋。

在脑子里拟订了一下：

跋

我讲三句话，刚才那句是第一句，谢谢。

7. 发现上面的悖论其实与“我只给不给自己刮胡子的男人刮胡子”一样，怀疑它是一个集合上的相互嵌套的东西。但脑子是模糊思维，现在还搞不清楚悖论的数学背景，或者没有时间搞清楚。

8. 发现自己的荨麻疹在断药后再次发作。于是决定上网查一下这个病情。后来觉得暂时不看西医。小心不吃到辣椒。可惜学校食堂的菜有很多辣椒。

9. 发短消息聊天，谈论冷雨夜我不想回家。

10. 看了40分钟的纤维丛。不太理解为什么一个圈上的2个v-bundle分别是柱面与mobius带。

11. pt送来的几张数码照片，偷拍了因为Sars在我们西西楼门口站岗的那个女孩子。

12. q的计划。（保密）

13. 又想起《寻秦记》，打算写一《寻商记》。问题是怎么样写，写多少，写了有什么意义。

14. 思考RN黑洞的裸奇点，不知道想什么。

15. 每天晚上听歌才能睡着，比如《囚鸟》。发现自己喜欢幽怨。
16. 楼前绿树郁郁葱葱，突然在树上挂了红色的宣传条。
17. Sars继续着，每天觉得日子在须臾间过去，在宿舍里老模仿ww的小孩子语言：慌了吧，虾米了吧~~~~。
18. 再看了《大话西游》一遍。觉得还是那么经典。周星驰是一个天才。
19. 为未来担心。
20. 看见达夫（duff），卢建新那本超弦的书的封皮。
21. 决定过小日子。
22. 想知道j, h, 很多人现在怎么样了。
23. 突然觉得自己失去了很多东西，比如每天与人擦肩而过。
24. 认为只有数学中的微分几何与拓扑学可以给自己带来快乐。
25. 还搞不清楚黎曼张量为什么有那几个独立分量。
26. 看费曼的物理学讲义几分钟，觉得太简单了。
27. 看狄拉克的量子力学，发现出版在1932年？又发现自己一点也不喜欢微扰论。到现在还是不懂为什么有选择定则。
28. 思考以后要不要考理论物理的研究生。认为研究生是一种趋势。
29. 发现黄祖洽先生虽然是院士，但他上的粒子物理我还不是很明白。
30. 打算回忆一下自己在3个星期的时间里到底做了什么。

2003年，正是我自己在学习量子力学和广义相对论的那一年，那一年的多愁善感和励精图治，已经成为过去，那一年冬天北师大的树上站满乌鸦，下的雪织成一张稠密的罗网，我没有感觉寒冷孤寂和疏离，因为我正在学习量子力学和广义相对论，我的那颗多愁善感的文人的灵魂得到了救赎。物理学本身不能让人温暖，但对一个曾经写小说的人来说，物理学会让我不再空虚。

萨特说：“存在就是虚无。他人就是地狱。”

卢梭说：“人生而自由，却无时不在枷锁之中。”

这种感慨不仅仅是书斋里的无病呻吟，而是那些文人哲学家为寻找

意义而痛哭流涕的独白。

我想，这本书对现在的年轻人来说，也会起到同样的效果，这本书不会让你懂量子力学与相对论，但可能会让你知道你自己不懂量子力学与相对论。相信读这本书的时候，你不会觉得空虚，因为物理里面，埋藏着永恒的东西。

张轩中

2013年10月

理解科学丛书·卢昌海科普著作

SEARCH FOR THE EDGE OF
THE SOLAR SYSTEM

那颗星星不在星图上

寻找太阳系的疆界

卢昌海◎著

天王星早在1690年就被记录过，为何直到1781年才被发现？

“海王星档案”隐藏了重大秘密吗？

冥王星缘何会被降级？

遥远的奥尔特云天体有可能被观测到吗？

太阳会有伴星吗？

清华大学出版社

目 录

[序](#)

[自序](#)

[引言](#)

[1 远古苍穹](#)

[2 乐师星匠](#)

[3 巡天偶得](#)

[4 命运弄人](#)

[5 虚席以待](#)

[6 失而复得](#)

[7 名分之争](#)

[8 轨道拉锯](#)

[9 众说纷纭](#)

[10 数学难题](#)

[11 星探出击](#)

[12 三访艾里](#)

[13 殊途同归](#)

[14 剑桥梦碎](#)

[15 欲迎还拒](#)

[16 生日之夜](#)

[17 名动天下](#)

[18 轩然大波](#)

[19 握手言和](#)

[20 秘密档案](#)

[21 先入之见](#)

[22 火神疑踪](#)

[23 无中生有](#)

[24 歧途苦旅](#)

[25 农家少年](#)

[26 寒夜暗影](#)

[27 大小之谜](#)

[28 深空隐秘](#)

[29 巅峰之战](#)

[30 玄冰世界](#)

[31 冥王退位](#)

[32 疆界何方](#)

[附录 冥王星沉浮记](#)

[术语表](#)

[人名索引](#)

[术语索引](#)

[参考文献](#)

[返回总目录](#)

作者简介

卢昌海，出生于杭州，本科就读于复旦大学物理系，毕业后赴美留学，于2000年获美国哥伦比亚大学物理学博士学位，目前旅居纽约。著有《从奇点到虫洞：广义相对论专题选讲》、《太阳的故事》和《黎曼猜想漫谈》，并曾在《中国青年报》、《数学文化》、《科幻世界》、《现代物理知识》、《中学生天地》、《科学画报》等报纸、杂志上发表几十篇科普及专业科普作品。

SEARCH

FOR THE EDGE OF THE SOLAR SYSTEM

那颗星星不在星图上：寻找太阳系的疆界

卢昌海 著

清华大学出版社
北 京

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

那颗星星不在星图上：寻找太阳系的疆界 / 卢昌海著. —北京：清华大学出版社，2013

（理解科学丛书）

ISBN 978-7-302-33821-5

I. ①那... II. ①卢... III. ①太阳系—普及读物 IV. ①P18-49

中国版本图书馆CIP数据核字（2013）第212282号

责任编辑：邹开颜

封面设计：蔡小波

插图：李璟

责任校对：刘玉霞

责任印制：沈露

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社总机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者：三河市金元印装有限公司

经 销：全国新华书店

开 本：165mm×240mm

印 张：12.5

字 数：172千字

版 次：2013年12月第1版

印 次：2013年12月第1次印刷

产品编号：051549-01

序

SEARCH

FOR THE EDGE OF THE SOLAR SYSTEM

我与本书的作者是熟悉的。当年，我为复旦物理系高年级少数优秀学生开了一个讨论班，学习量子理论初期发展的历史，希望能够更好地理解其中的一些难点问题。就是在这个讨论班上，当时还是大学一年级新生的卢昌海，主动请求作个报告，要介绍海森堡的矩阵力学。可以想象，我当然是带着极其怀疑的眼光答应了他的请求，主要还是不想伤害一个年轻人的热情和自尊。但结果着实让我和他的学长们大吃一惊，他真的已经完全掌握这部分内容了！一年之后，他又提出要免修物理系最重头的全部“四大力学”课程，即理论力学、热力学与统计物理、量子力学和电动力学。为此，系里专门为他组成阵容超豪华的名教授团队，一门门笔试加口试地进行。全部结束之后，每一位参加测试的教授都真的被这个年轻人的才华折服了。据我所知，一位低年级学生能免修全部的“四大力学”课程，在复旦物理系的历史上还从未有过，而且成绩还是无可争辩的全优。或许这一“光辉纪录”还会保持相当长的时间吧。

就是这样一位当年的才子，今天已成为一位优秀的科普作家。除了这本新版的《那颗星星不在星图上——寻找太阳系的疆界》，清华大学出版社还出版了他的另外两部科普著作《太阳的故事》和《黎曼猜想漫谈》，都很精彩。其中，后一本书还得到了大数学家王元的褒奖和推荐。另外，昌海目前还在努力地写作，相信会有更多的佳作问世。

回想当年，一套《十万个为什么》几乎成为我们这代人青少年时期科普作品的代名词。所幸的是，这种时代一去不复返了。今天的情景已完全不同了，书店里的科普作品可谓琳琅满目。多是多矣，然而拿起来

翻阅几页后，还能不让人失望的却不多见。归纳起来可以说，一些作者对什么是真正好的科普作品还缺乏认识。第一，科普作品绝非“浅”知识的堆积，更不是一堆知识，知识一堆。第二，科普作品需要将深奥的道理和知识用浅显的语言讲出来，道明白，但它不应该被庸俗化，更不允许被误导。第三，如果科普作品的文字（包括翻译的文字），读起来比作品内容本身还难懂的话，怎能不让人沮丧而无语呢？

事实上，若非才、学、识皆备，很难写出好的科普作品。昌海的这本书就是这样一本难得的佳作，这是一次从地球出发的太空“深度游”。作者的“才”就在于他能将那些重要“景点”的来龙去脉交代得清清楚楚，如数家珍，让人有身临其境之感。在不知不觉、轻松愉快的气氛中，对太阳系的结构形成了一幅生动的物理图像。有别于一般专业作品，一部好的科普作品，要求作者有好的文字。昌海的文字表达不仅简洁、干净，而且还能在一些节骨眼上展现幽默和诙谐，读起来赏心悦目。作者的“学”体现在对那些常被人讹传或误解、夸张的历史事件进行分析和澄清，证据确凿，令人信服。在逐字逐句地通读完这本书之后，最令我佩服的是作者的“识”，也就是他对物理或说对科学的品味。对于时间跨度如此之长，空间上如此遥远而又神秘莫测的有关太阳系边界的探索之路，在这样一本小书中得到如此惊心动魄而又深入浅出的刻画，如果没有好的品味，完全没有可能做到。

俗话说，好东西应该与好朋友分享。昌海的这本书，在我身边的朋友中已有相当大的“知名度”了，但那只不过是几个人而已。正是考虑到这一因素，当清华大学出版社邀请我为新版的书作序时，我欣然答应，而且可以很肯定地说，每个拿起这本书翻阅的人一定不会失望。

金晓峰
2013年5月

自序

SEARCH

FOR THE EDGE OF THE SOLAR SYSTEM

在我为自己的第三本书《黎曼猜想漫谈》撰写后记时，曾对前两本书没有前言或后记的原因作过这样的解释：

并不是不想写，而是因为那两本书的写作及出版过程都很平淡（或曰顺利），没什么值得叙述的。若生添一篇前言或后记，不免有灌水之嫌。

现在，我却要为那两本书中的第一本——《寻找太阳系的疆界》的修订版“生添”一篇自序了，其“灌水之嫌”且容我辩白几句（希望不会越辩越黑）。

之所以要写这篇自序，主要有两个原因。首先是因为距离本书初版的问世已经过了三年多，在如今这个快节奏的时代里，算是一段不太短的时间了。而且对于本书来说，这三年多的时间颇具代表性，甚至可以说是走过了一个生死轮回，从而多少有了一点谈“历史”的资历——就像久历了岁月的人多少可以谈点往事一样。

其次是因为修订版——或许是出于促销方面的考虑——对书名作了变更。我虽由衷地希望出版社不要因出版我的作品而亏损，心底里却更害怕读者因书名变更而将修订版当成新书误买以致血压升高，因此想在尽可能靠前的文字——即这篇自序——中提个醒。不过，这一提醒是否真有效力却殊难预料，因为读者买书前未必都会看自序，网购的读者则是想看也未必看得到。倘若哪位读者不幸仍中了书名变更之“招”，致使

足可购买若干个汉堡包的私款流失，可到我的网站

（<http://www.changhai.org/>）来留言解恨。

好了，现在言归正传，谈点与本书有关的往事吧。本书的撰写始于2007年3月，一开始只是作为系列文章在我的网站上连载。连载了几篇之后，恰逢杭州《中学生天地》杂志的一位编辑来信约稿，我便提及了该系列，编辑看后表示有兴趣。于是自2007年9月起，本书的内容开始在《中学生天地》杂志上连载。不过，由于杂志方面对字数有一定的限制，因此刊出的往往是删节版，尤其是到了后期，杂志方面希望在一年之内完成连载，比我自己对内容的规划少了好几个月，因此最后几期刊出的内容存在大幅度的删节。但另一方面，杂志的连载虽有诸多欠缺，却正是由于要向杂志供稿，使那个系列成为我撰写的篇幅相近的所有系列中最先完成的。从这点上讲，杂志的连载功不可没。《寻找太阳系的疆界》的单行本于2009年11月出版，成为我的第一本书，也在一定程度上得益于此。

不过，《寻找太阳系的疆界》的写作及出版过程虽然顺利，出版后的命运却不无曲折。初版的问世才不过三年，就陷入了极大的窘境，其结果用我网站上一位网友的话说，是成为了“绝版名著”。当然，那是戏言——确切地说，后两个字（“名著”）是戏言（虽然我很希望不是戏言），前两个字（“绝版”）却是事实（虽然我很希望不是事实），因为本书的初版确实已无处购买了（除非是购买旧书）。只不过那并非因为卖得太好以致脱销，而恰恰相反，乃是因为卖得太不好，以至于未及卖完，就被清了库存。对图书来说，可以说是“死”了一回。

唯一值得庆幸的，是本书的零售虽十分失败，却“东边不亮西边亮”地中标了若干个省份的中小学图书的馆配，从而成为了一些中小学生的“钦定”课外读物之一。也许是这个缘故，出版社决定为本书再冒一次险，出一个修订版。本书因此而有了如今这个“死而复生”的机会。

那么，这个所谓修订版究竟在何处作了修订呢？从正文上讲，只是

更正了几处笔误，并扩充了几个注释，可以说是微乎其微的（这是托“历史题材”之福，因为科学史不像科学前沿那样日新月异）。不过，图书的修订并不限于正文，本书的真正修订是以下三类内容：

（1）插图——修订版添加了许多新插图，而且是手工绘制的，不同于初版中那些来自互联网的现成图片。

（2）索引——包括人名和术语两部分，索引在国外科普图书中几乎已是必不可少的组成部分，在国内科普图书中却还不太普遍，在我自己的作品中则是首次添加。

（3）文字——包括序言（由复旦大学物理系的金晓峰老师所撰）、附录（由我2009年10月以删节版形式发表在《科学画报》上的《冥王星沉浮记》一文的完整版整理而成）及自序（即本文）。

以上就是对本书及修订版的简单介绍。说实话，对于出版社此次的“冒险行动”我是暗暗捏一把汗的。作为作者，我对自己作品的水准是有信心的，但作为有几十年读书、买书经验的资深书迷，我却深知那绝不等于能卖得好。玩过博客的朋友们大都知道，非著名作者在非热门话题上哪怕写上十篇“沥血之作”，也赶不上知名人士贴一张宠物相片更有点击数。这是大众行为的鲜明特点，非独博文如此。不过，在捏汗的同时，我还是要感谢清华大学出版社的“冒险”，并且特别感谢为本书及修订版的出版付出巨大心力的邹开颜编辑（她也是我其他几本书的编辑）。另外，我也要感谢为本书修订版撰写序言的金晓峰老师，在平面媒体或博客上为本书初版撰写过书评的秦克诚、陈学雷等先生，为本书绘制插图的李璟小姐，以及本书过去、现在和将来的所有读者。

引 言

记得念小学的时候，读过一篇课文，叫做“数星星的孩子”，讲述汉朝天文学家张衡的童年故事。时隔这么多年，小学的很多课文我已经忘记了，但那篇数星星的课文却依然历历在目。那时候，我住在杭州的郊外，家门口有一个池塘，在许多个晴朗的夏夜里，我和小伙伴们也常常坐在池塘边仰望星空。那时候，郊外的天空还没有被都市的灯光所污染，在广袤的天幕下，那一颗颗璀璨夺目的星星显得格外的晶莹和美丽。自远古以来，这种无与伦比的美丽就吸引了一代又一代的追随者，他们中的一些人甚至将自己的一生都献给了探索星空奥秘的科学事业。人类寻找太阳系疆界的故事只是科学史上的几朵小小浪花，但在那些故事中，有浪漫，也有艰辛，有情理之中，也有意料之外，有功成名就的兴奋，也有错失良机的遗憾，它们就像天上的星星一样美丽动人。



1 远古苍穹

很多故事都会用“很久很久以前”作为开始，仿佛久远的年代是成就一个好故事的要素。现在让我们也从“很久很久以前”开始，来讲述人类寻找太阳系疆界的故事吧。

在很久很久以前，一群古希腊的牧羊人孤单地生活在辽阔的原野上。他们白天与羊群为伍，在原野上漫游，夜晚则与星空为伴，期待黎明的到来。渐渐地，他们注意到在黎明之前，在晨光渐露、太阳即将跃出地平线的时候，天边有时会出现一颗闪烁的星星。与多数星星不同的是，那颗星星的位置会一天天地变化，有时甚至会连续一段时间不出现。他们把这颗出现在黎明时分的星星叫做“晨星”（morning star）。细心的牧羊人还注意到，在黄昏时分，在日沉大地、暮色四合的时候，天边有时也会出现一颗闪烁的星星，它的位置也会一天天地变化，有时也会连续一段时间不出现。他们把那颗出现在黄昏时分的星星叫做“晚星”（evening star）。后来人们用希腊及罗马神话中的太阳神阿波罗（Apollo）表示晨星，用希腊或罗马神话中的信使赫耳墨斯（Hermes）或墨丘利（Mercury）表示晚星。很多年之后，人们意识到晨星和晚星实际上是出现在不同时刻的同一颗星星，据说毕达哥拉斯

（Pythagoras）是最早意识到这一点的人^{[\(1\)](#)}。在群星之中，这颗星星的位置变化最为显著，往来如梭，仿佛天空中的信使，信使墨丘利便成了它的名字。

像这样的小故事在人类文明的几乎每一个早期发源地都曾有过。那时的人们就已经知道，在浩瀚的夜空中，多数星星的位置看上去是固定的，像晨星（晚星）这样会移动的星星是十分少见的。这样的星星被称为行星，它的英文名planet来自希腊文πλανήτης（planētēs），其含义是漫游者。远古人类所发现的行星共有五颗。这个数目在长达几千年的时

间里从未改变过，甚至一度被认为是永恒不变的真理。在东方的中国及深受中华文化影响的其他东方国家如日本、韩国及越南，人们将五颗行星与阴阳五行联系在一起，并以此将它们分别命名为水星（即上面提到的墨丘利（Mercury）），金星（在西方世界中被称为维纳斯

（Venus），她是罗马神话中掌管爱情与美丽的女神），火星（在西方世界中被称为玛尔斯（Mars），他是罗马神话中的战神），木星（在西方世界中被称为朱庇特（Jupiter），他是罗马神话中的众神之王）和土星（在西方世界中被称为萨坦（Saturn），他是朱庇特的父亲，是罗马神话中掌管农业与收获的神）。很明显，这种命名方式除了起到命名作用外，还代表了古代东方文化对行星数目“五”的一种神秘主义的解读。类似的解读方式不仅存在于东方，也存在于西方；不仅存在于古代，也存在于近代。哥白尼（Nicolaus Copernicus）的日心说提出之后，地球本身也被贬为了行星，行星的数目由“五”变成了“六”。对此，著名的德国天文学家开普勒（Johannes Kepler）提出了一个几何模型（图1），试图将天空中存在六颗行星与三维空间中存在五种正多面体这一几何规律联系在一起^{[\(2\)](#)}。

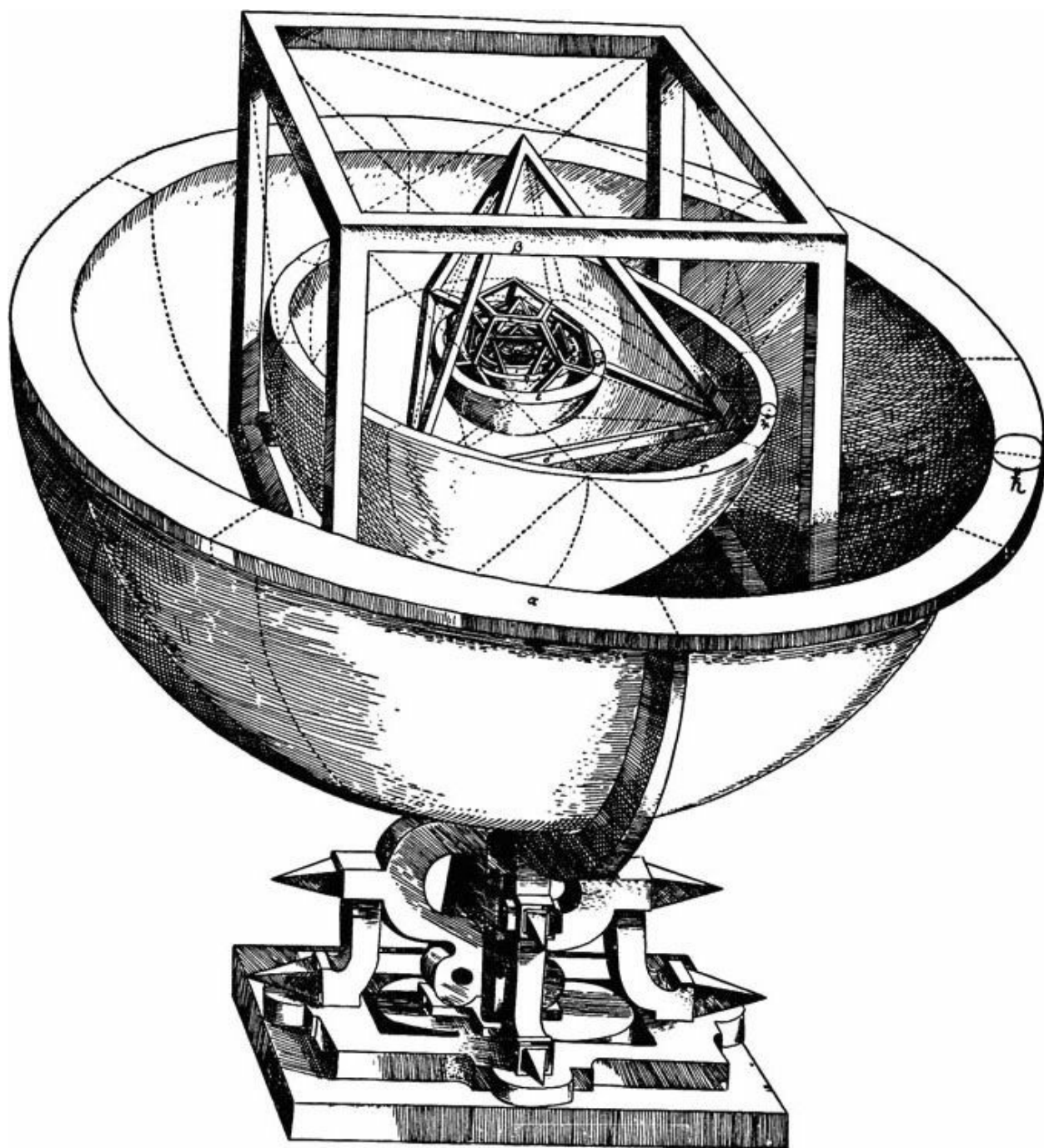


图1 开普勒的行星模型

诸如此类的对行星数目的神秘主义解读虽然并没有什么生命力，但除了因日心说导致的地球地位变更外，行星数目的长期不变却是不争的事实。一百年、两百年……一千年、两千年……，这个数目是如此的根深蒂固，天文学家们大都将之视为不言而喻的事实了。他们也许做梦也

没想到，这个数目有一天竟然也会改变。这一天是1781年3月13日，改变这个数目的是生活在一座英国小镇的一位业余天文学家，他的名字叫做赫歇耳（William Herschel）。他发现了太阳系的第七颗行星，从而成为几千年来发现新行星的第一人。赫歇耳的发现出乎了包括他自己在内的所有人的意料，这一发现不仅为他本人赢得了永久的荣誉，也将观测天文学带入了一个崭新的时代，一个由赫歇耳“无心插柳”而开启的天文学家们“有心栽花”的时代，人类从此开始了寻找太阳系疆界的漫漫征途。



注释

[\[1\]](#) 除墨丘利（即水星）外，另一颗内行星——金星——也只有在清晨和黄昏才容易被肉眼所看见（请读者想一想，为什么水星和金星只有在清晨和黄昏才容易被肉眼所看见？），因而也曾被远古的观测者误分成晨星和晚星。后来也是古希腊人首先意识到它们其实是出现在不同时刻的同一颗行星。

[\[2\]](#) 具体地讲，开普勒提出的几何模型是这样的：将六颗行星与三维空间中仅有的五种正多面体按以下顺序自内向外排列：水星、正八面体、金星、正二十面体、地球、正十二面体、火星、正四面体、木星、正六面体、土星。排列的方式是：每个行星轨道所在的球面都与其外侧的正多面体相内切（最外侧的土星轨道除外），同时与其内侧的正多面体相外接（最内侧的水星轨道除外）。开普勒的这一模型虽然精巧，但与精密的观测以及他自己后来发现的行星运动定律不相符合，不久之后就被放弃了。喜欢几何的读者不妨计算一下这一模型所给出的相邻行星的轨道半径之比，并与观测数值作一个比较。

2 乐师星匠

赫歇耳的一生非常出色地实践了两种截然不同的职业，其中最出色的职业——天文学家——不仅出现在对常人来说很难有开创性成就的后半生里，而且从某种意义上讲，就像他对新行星的发现一样，是一个无心插柳的故事。



英国天文学家 赫歇耳（1738-1822）

赫歇耳于1738年11月15日出生在当时属于英王领地的德国中北部城市汉诺威（Hanover）的一个音乐之家⁽¹⁾。赫歇耳具有很高的音乐天赋，他14岁就参加乐队，不仅擅长多种乐器，而且还能独立作曲，他亲自创作的交响曲和协奏曲就有几十首之多。1757年秋天，19岁的赫歇耳移居到了英国⁽²⁾，以演奏及讲授音乐为生。

赫歇耳的音乐成就以常人的标准来衡量应该说是颇为可观的，但放在他的简历中，却无可避免地要被他巨大的天文成就所淹没。不过他在英国的音乐生涯中有一件事情值得一提。那是在18世纪60年代中期，当时英国的教会刚刚开始引进风琴，需要招募一批风琴演奏者，年轻的赫歇耳也参加了一个风琴演奏职位的竞逐。当时的竞争颇为激烈，而赫歇耳在风琴演奏上并无经验。但他敏锐地发现当时英国教会引进的风琴与欧洲大陆的风琴相比有一个缺陷，那就是缺少控制低音部的踏板。为了弥补这一缺陷，聪明的赫歇耳对两个低音琴键进行了改动，从而演奏出了通常需要低音踏板的配合才能演奏出的低音部。他的表演不仅赢得了评审的一致赞赏，而且让他们深感神秘（当然，他顺理成章地成为了优胜者）。赫歇耳在这一竞争中显示出过人的动手及设计能力，将为他日后的天文生涯立下汗马功劳。

1766年，赫歇耳迁居到了英国西南部的一座名叫巴斯（Bath）的小镇，在一所教堂担任风琴演奏师，开始了他在那里长达16年的生活（图2）。这座当时人口仅有两千的观光小镇因而有幸见证了赫歇耳一生最辉煌的工作。在巴斯期间，赫歇耳的音乐生涯达到了巅峰，他不仅是风琴演奏师，而且还担任了当地音乐会的总监，并开班讲授音乐课程。1772年，收入已颇为殷实的赫歇耳给他母亲寄去了足够雇一位佣人的钱，从而把他妹妹夏洛琳（Caroline Herschel）从母亲为她安排的枯燥繁重的家务劳动中解救了出来，并接到巴斯。



图2 赫歇耳位于巴斯的住所（已辟为博物馆）

与赫歇耳一样，卡洛琳也是一位颇有音乐天赋的人，但她一生注定要跟随哥哥去走一条未曾规划过的道路。在接卡洛琳到巴斯之前，已成为镇上知名音乐家的赫歇耳潜心学起了数学。赫歇耳学数学的本意是想多了解一些和声的数学机理，从而加强自己的音乐素养。但结果却因学数学而接触了光学，又因接触光学而对天文学产生了浓厚的兴趣，最终走上了一条业余天文学家之路。而卡洛琳则成为了他在天文观测上不可或缺的助手^[3]。

赫歇耳所走的这条业余天文学家之路，不仅为他自己走出了一片绚烂的天地，也成就了业余天文学的一段——也许是最后一段——黄金岁月。18世纪的许多职业天文学家过分沉醉于由牛顿（Isaac Newton）所奠定，并经欧拉（Leonhard Euler）、拉格朗日（Joseph Louis

Lagrange）、拉普拉斯（Pierre Simon Laplace）等人所改进的辉煌的力学体系之中。他们热衷于计算各种已知天体的轨道，以此检验牛顿力学，同时也为经纬及时间的确定提供精密参照。在一定程度上，当时的许多职业天文学家变得精于验证性的计算，却疏于探索性的观测。在这种情况下，自赫歇耳之后半个多世纪的时间里，业余天文学家们对天文学的发展起了重要的补充作用，这一时期天文学上的许多重大的观测发现就出自他们之手。

常言道：“工欲善其事，必先利其器。”对天文观测来说，必备的工具是望远镜。由于当时高质量的望远镜极其昂贵，赫歇耳决定自己动手制作望远镜（也顺便可以实践因学数学而接触的光学知识）。望远镜的问世是在17世纪初，其确切的发明者现已无从追溯，但德国裔荷兰人利普歇（Hans Lippershey）于1608年最早为自己制作的望远镜申请了专利，从而留下了文字记录，因此人们一般将他视为望远镜的发明者。

1609年，科学巨匠伽利略（Galileo Galilei）在得知了有关望远镜的消息后，很快制作出了自己的望远镜。伽利略制作的望远镜在结构及放大率上都大大优于包括利普歇在内的同时代人制作的望远镜。并且他也是最早将望远镜用于天文观测的人⁽⁴⁾。通过望远镜，伽利略获得了一系列前所未有的天文发现，其中包括发现月球上的环形山、太阳黑子及木星的四颗卫星（现在被称为伽利略卫星）等。不过伽利略所用的是折射望远镜，这种望远镜由于透镜（主要是物镜）所具有的色差等当时技术难以消除的效应而无法达到很高的放大率。17世纪后期，另一位科学巨匠牛顿发明了反射望远镜⁽⁵⁾，用反射面替代了折射望远镜中的物镜，从而避免了透镜色差带来的困扰。赫歇耳所制作的就反射望远镜，这种望远镜的反射面可以用金属制作而无需使用玻璃。

为了制作望远镜，赫歇耳将自己在巴斯的住所改造成了望远镜“梦工厂”：客厅被用来制作镜架与镜筒，卧室变成了研磨目镜的场所，厨房里则架起了熊熊的熔炉。赫歇耳细心试验了许多不同成分的合金，最

后选择了用71%的铜与29%的锡组成的合金，作为制作反射面的材料。在制作望远镜期间，除妹妹卡洛琳外，赫歇耳还得到了弟弟亚历山大（Alexander Herschel）的帮助。赫歇耳一生制作的望远镜有几百架之多，不仅满足了自己的需要，而且还通过出售望远镜使家庭获得了数目不菲的额外收入。在长期的制作中，他的作坊也一度发生过严重事故，导致熔融的金属四处飞溅，幸好大家闪避及时，奇迹般地未造成人员伤亡。

1778年，赫歇耳的家庭作坊制作出了一架直径6.2英寸、焦距7英尺的反射望远镜。这架望远镜在天文史上有着重要的意义，被后世称为“七英尺望远镜”（图3）。后来的检验表明，赫歇耳这架“七英尺望远镜”的性能全面超越了当时英国格林威治（Greenwich）皇家天文台的望远镜。赫歇耳用自己的双手制造出了当时全世界最顶尖的观测设备，为自己的天文观测之路迈出了无比坚实的第一步。终其一生，赫歇耳孜孜不倦地建造着更大的望远镜，一次再次地刷新着自己——从而也是整个天文学界——的纪录，他在这一领域的优势不仅在其有生之年从未被反超过，甚至在去世之后仍保持了很长时间。

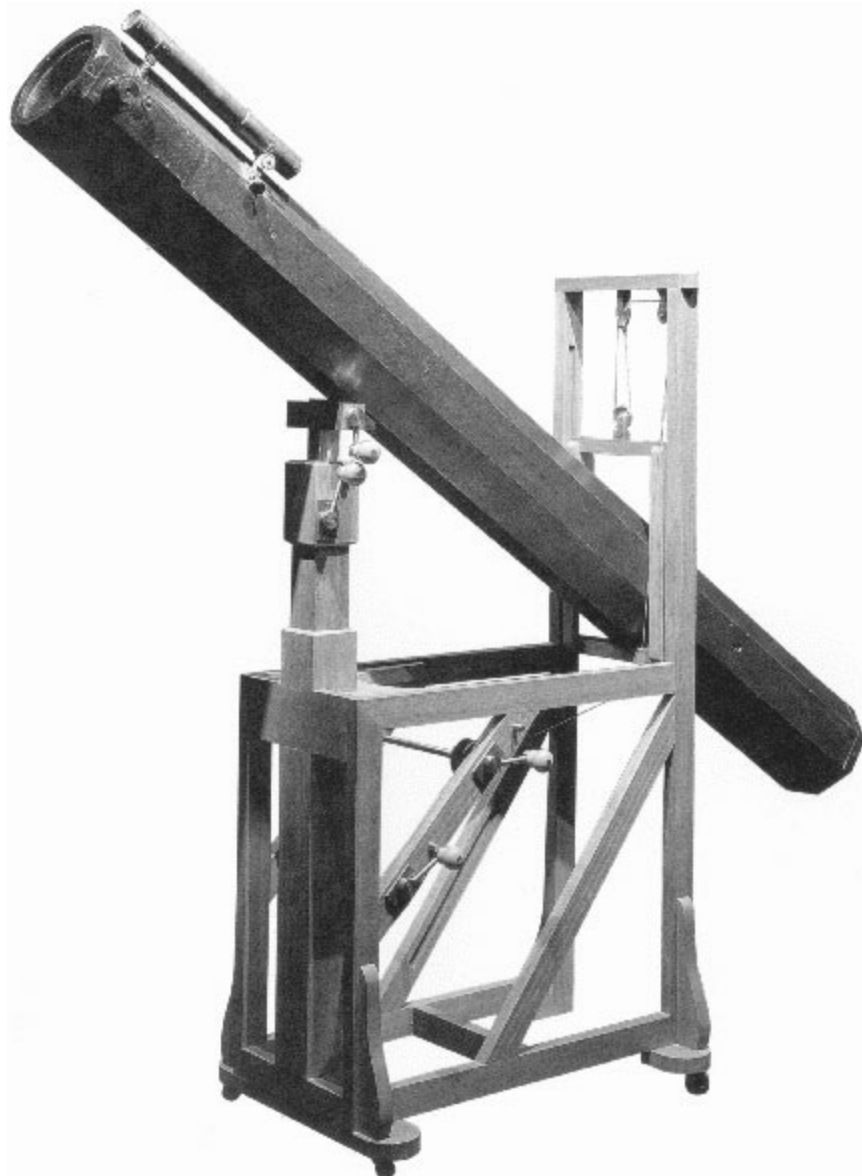


图3 赫歇耳的“七英尺望远镜”

三年后的一个春季的夜晚，一颗略带圆面的星星出现在了赫歇耳那架“七英尺望远镜”的视野里，他一生最伟大的发现来临了。

注释

[\[1\]](#) 赫歇耳出生时的名字是Friedrich Wilhelm Herschel，后来所用的名字Frederick William Herschel是他移居英国后入乡随俗而改的。确切地讲，为了与后文用卡洛琳（Caroline）称呼他

妹妹Caroline Herschel，以及用亚历山大（Alexander）称呼他弟弟Alexander Herschel相平行，我们应该称他为威廉（William）。不过由于他是科学史上的著名人物，对这样的人物，人们习惯于用姓而不是名来称呼，就像我们一般不用艾萨克（Isaac）和阿尔伯特（Albert）来称呼牛顿（Isaac Newton）和爱因斯坦（Albert Einstein）一样。

[\[2\]](#) 在此之前，赫歇耳曾在英国逗留过大约9个月，较好地掌握了英语。

[\[3\]](#) 卡洛琳自己后来也成为了一位天文学家，她在寻找彗星方面有不俗的成就，总共发现了八颗彗星。

[\[4\]](#) 值得注意的是，伽利略在其早期著作《星际使者》（The Starry Messenger）的开篇曾以第三人称的口吻将望远镜说成是自己的发明（不过他在正文中提到自己在制作望远镜之前听说过他人制作望远镜的消息）。由于这段文字的影响，伽利略曾被一些人视为是望远镜的发明者，这一说法如今已被否定。不过平心而论，伽利略在改进望远镜方面所做的贡献是巨大的，不仅大大提高了放大率，而且据说是他首先解决了望远镜成像的上下倒置问题。另外，他在制作自己的望远镜之前只是听说过有关望远镜的消息，而未见过实物。因此将伽利略视为望远镜的发明者之一也并不过分。

[\[5\]](#) 反射望远镜的设计在牛顿之前就已存在，但牛顿最早制作出了具有实用价值的反射望远镜。牛顿的制作水平之高，使伦敦的工匠们在几年之后都没有能力加以效仿。

3 巡天偶得

天文观测在外人看来也许是一项很浪漫的事业，但实际上虽不乏浪漫，却也充满了艰辛。即便拥有高质量的望远镜，一项天文发现的背后也往往凝聚着天文学家长年累月的心血。赫歇耳不仅在制作望远镜上走在了同时代人的前面，在天文观测上也有着常人难以企及的细心和热忱。他一生仅巡天观测就进行了四次之多，每一次都对观测到的天体进行了系统而全面的记录。其中最早的一次是通过一架口径4.5英寸的反射望远镜进行的，涵盖的是所有视星等亮于4的天体⁽¹⁾。由于视星等亮于4的天体用肉眼都清晰可见，这样的观测对于他精心制作的望远镜来说无疑只是牛刀小试。而且，这类天体既然用肉眼就能看见，从中做出任何重大发现的可能性显然都是微乎其微的。用功利的眼光来看，这样的巡天观测几乎是在浪费时间，但对赫歇耳来说，天文观测的乐趣远远超越了任何功利的目的。从这样一次注定不可能有重大发现的巡天观测开始自己的观测生涯，极好地体现了赫歇耳在天文观测上扎实、沉稳、严谨、系统的风格。除了这种极具专业色彩的风格外，赫歇耳对天文观测的酷爱程度也是非常罕见的。他对观测的沉醉，实已达到了废寝忘食的境界。在他从事观测时，食物常常是卡洛琳用勺子一小口一小口地喂进他的嘴里，而睡觉则往往要托坏天气的福。正是这样的专业风格与忘我热忱的完美结合，最终成就了天文观测史上的一次伟大发现。

几年下来，赫歇耳以及他所制造的望远镜在英国学术圈里渐渐有了一些知名度。“七英尺望远镜”制作完成后，赫歇耳开始用这架举世无双的望远镜进行自己的第二次巡天观测，这次巡天观测的目的之一是寻找双星（赫歇耳一生共找到过800多对双星，是研究双星的前驱者之一），所涵盖的最暗天体的表观亮度约为8等，相当于上次巡天观测所涉及的最暗天体表观亮度的1/40，或肉眼所能看到的最暗天体表观亮度

的1/6。显然，这次巡天观测所涉及的天体数量比上一次大得多，工作量也大得多。

1781年3月13日夜晚10点到11点之间，赫歇耳的望远镜指向了位于金牛座（Taurus）一“角”（ζ星）与双子座（Gemini）一“脚”（η星）之间的一小片天区。在望远镜的视野里，一个视星等在6左右，略带圆面的新天体引起了赫歇耳的注意。那会是一个什么天体呢？由于恒星是不会在望远镜里留下圆面的，因此这一天体不像是恒星。为了证实这一点，赫歇耳更换了望远镜的镜片，将放大倍率由巡天观测所用的227倍增加到460倍，尔后又进一步增加到932倍，结果发现这个天体的线度按比例地放大了。（请读者思考一下，赫歇耳既然有放大率更高的镜片，在巡天观测时为什么不用？）毫无疑问，这样的天体绝不可能是恒星，恒星哪怕在更大的放大倍率下也只会是一个亮点，而不会呈现出圆面。那么，它究竟是一个什么天体呢？赫歇耳认为答案有可能是星云状物体，也有可能是彗星。但就在他试图一探究竟的时候，巴斯的天公却不作美，一连几天都不适合天文观测，赫歇耳苦等了四天才等来了再次观测这一天体的机会，这时他发现该天体的位置与四天前的记录相比，有了细微的移动。由于星云状物体和恒星一样是不运动的，因此这一发现排除了该天体为星云状物体的可能性。于是赫歇耳的选项只剩下一个，那就是彗星，他正式宣布自己发现了一颗新的“彗星”。



发现新彗星虽然算不上是很重大的天文发现，但每颗新彗星的发现都能为天文学家们新增一个研究轨道的对象，而这在当时正是很多人感兴趣的事情。因此天文学家们一得知赫歇耳发现新“彗星”的消息，便立即对新“彗星”展开了观测。令人奇怪的是，这颗新“彗星”并没有像其他彗星那样拖着长长的尾巴。用后人的眼光来看，或许很难理解如此显著的疑点为何没有让赫歇耳意识到自己所发现的其实不是彗星，而是一颗新的行星。但在当时，“新行星”这一概念对很多人来说几乎是一个思维上的盲点。不过科学家毕竟是科学家，他们是不大会始终沉陷在盲点里漠视证据的。赫歇耳的发现公布之后，英国皇家学会的天文学家马斯克林（Nevil Maskelyne）在对该“彗星”进行了几个夜晚的跟踪观测之后，率先猜测它有可能是一颗新的行星，因为它不仅没有彗星的尾巴，连轨道也迥异于彗星。当然，凭借短短几个夜晚的观测，马斯克林只能对新天体的轨道进行很粗略的推断。几个月之后，随着观测数据的积累，瑞典天文学家莱克塞尔（Anders Johan Lexell）、法国科学家萨隆（Bochart de Saron），以及法国天体力学大师拉普拉斯彼此独立地从数学上论证了新天体的轨道接近于圆形，从而与接近抛物线的彗星轨道截然不同。与此同时，赫歇耳本人也借助自己无与伦比的望远镜优势对新天体的大小进行了估计，结果发现其直径约为54700千米，是地球直径的4倍多^[2]。显然，在近圆形轨道上运动的如此巨大的天体只能是行星，而绝不可能是彗星。因此到了1781年的秋天，天文学界已普遍认为赫歇耳发现的是太阳系的第七大行星。这颗行星比水星、金星、地球和火星都大得多，甚至比它们加在一起还要大得多，它绕太阳公转的轨道半径约为30亿千米，相当于土星轨道半径的两倍，或地球轨道半径的20倍。

几千年来，人类所认识的太阳系的疆界终于第一次得到了扩展^[3]。

赫歇耳的伟大发现立即被英国天文学界引为骄傲，赫歇耳本人也因

此而获得了巨大的荣誉。1781年11月，英国皇家学会将自己的最高奖——考普雷奖（Copley Medal）授予了赫歇耳，并接纳他为皇家天文学会的成员。赫歇耳从此成为了职业天文学家。为了让赫歇耳有充裕的财力从事研究，皇家学会免除了他的会费。不仅如此，英王乔治三世还特意为他提供了津贴，并亲自接见了他。后来乔治三世干脆请赫歇耳迁居到温莎堡（Windsor Castle）附近，以便能让他时常向皇室成员讲解星空知识。作为回报，赫歇耳在皇家学会的提示下写了一封感谢信，盛赞乔治三世对他的慷慨资助，并提议将新行星命名为“乔治星”（Georgian Planet）。虽然在新天体的命名中发现者通常享有优先权，但像“乔治星”这样一个富有政治意味的名字还是立即遭到了英国以外几乎所有天文学家的一致反对。赫歇耳本人也私下承认，这个名字是不可能被普遍接受的。在新行星的命名竞赛中最终胜出的，是德国天文学家波德（Johann Elert Bode），他提议的名称是乌拉诺斯（Uranus），这是希腊神话中的天空之神，也是萨坦（土星）的父亲。这一名称之所以胜出，是由于它与太阳系其他行星的命名方式具有明显的传承关系：在其他行星的命名中，朱庇特（木星）是玛尔斯（火星）的父亲，萨坦（土星）是朱庇特（木星）的父亲，有这样一连串“父子关系”为后盾，在土星之外的行星以萨坦（土星）的父亲乌拉诺斯来命名无疑是顺理成章的[\[4\]](#)。在中文中，这一行星被称为天王星。

发现天王星的那年赫歇耳已经42岁，一生的旅途已经走过了一半。在后半生里，他放弃了音乐生涯，将全部的精力都投注在了星空里，孜孜不倦地继续自己的天文事业，并且作出了卓越的贡献。除发现天王星外，他还分别发现了土星及天王星的两颗卫星[\[5\]](#)。他在恒星天文学、双星系统及银河系结构等领域的研究都具有奠基意义。他所绘制的星图远比以往的任何同类星图都更全面，同时他还是最早发现红外辐射的科学家。

1822年8月25日，赫歇耳在自己工作了几十年的观星楼里离开了人

世。他的一生只差3个月就满84岁，只差4个月就是他所发现的天王星绕太阳公转一圈的时间。

注释

[\[1\]](#) 视星等是描述天体表观亮度的参数，视星等越低，天体的表观亮度就越高。具体地讲，1等星的表观亮度是6等星的100倍。（请读者从中推算一下，视星等每降低1等，表观亮度会增加多少？）正常的肉眼在最佳观测条件下所能看到的最暗天体的视星等约为6等。

[\[2\]](#) 赫歇耳得到的这一数值略大于现代观测值，后者为赤道直径51118千米，两极直径49946千米。

[\[3\]](#) 这里我们没有把质量微不足道的彗星计算在内。

[\[4\]](#) 在新行星的命名基本得到公认之后，一些英国天文学家仍固执地延用着“乔治星”这一名称，直至19世纪中叶。

[\[5\]](#) 赫歇耳晚年曾认为自己还发现了天王星的另外四颗卫星，但那些“发现”后来要么被证实是错误的，要么因存在明显的疑点而未得到公认。

4 命运弄人

听完了发现天王星的故事，有读者也许会提出这样一个问题，那就是天王星为什么没有更早地被人们发现呢？我们前面提到过，天王星的视星等在6左右，事实上，在最亮时它的视星等甚至可以达到5.5。（请读者想一想，什么情况下天王星看起来会最亮？）这样的亮度连肉眼都有可能勉强看到，却为何没有更早地就被人们发现呢？赫歇耳之前的天文学家们虽然没有像“七英尺望远镜”那样出色的望远镜，但他们的望远镜用来观测像天王星这样一个原则上连肉眼都有可能看到的天体却是绰绰有余的。自伽利略之后的那么多年里，那么多的天文学家在那么多个晴朗的夜晚仰望苍穹，却为何会将发现新行星的伟大荣誉留到1781年呢？

我们在前面的叙述中已经知道，赫歇耳发现天王星的过程并不是一个有意寻找新行星的过程，甚至在发现天王星之后他还一度将之视为彗星。这一切都表明天王星的发现带有一定的偶然性，是一个“无心插柳”的过程。与赫歇耳同时代的一些天文学家曾因此而将赫歇耳对天王星的发现视为是好运气之下的偶然发现。赫歇耳的一生对荣誉大体是看得比较淡的，但他对这种将他发现天王星的过程视为偶然的说法还是明确表示了反对。他写下了这样的文字：

我对天空中的每颗星星都进行了有规律的排查，不仅包括（像天王星）那样亮度的，还包括许多暗淡得多的，它（天王星）只不过是恰好在那个夜晚被发现。我一直就在逐渐品读大自然所写的伟大著作，如今恰好读到了包含第七颗行星的那一页。假如有什么事情妨碍了那个夜晚，我必定会在下个夜晚发现它。我望远镜的高品质使得我一看到它便能感觉出它那可以分辨的行星圆面。

赫歇耳的这段文字不仅为自己发现天王星的必然性做了注解，而且也很好地说明了为什么在他之前那么多的天文学家都一直没有发现天王星。要知道，看到一颗暗淡的新行星虽然困难，但比这困难得多的则是要判断出它是行星而不是恒星。天王星被发现之后，人们对历史上的天文记录进行了重新排查，结果发现天王星在赫歇耳之前起码已被记录了22次之多，其中最早的一次可以追溯到1690年，比赫歇耳早了将近一个世纪。可惜留下这22次记录的天文学家们无一例外地与发现天王星的伟大荣誉擦肩而过。之所以会如此，是因为其中没有一位意识到自己观测到的不是恒星，而是行星。我们知道，在气象条件良好的夜晚，单凭肉眼就可以看到数以千计的星星，借助小型望远镜的帮助所能看到的天体数量更是多达数十万，这其中绝大多数都是恒星，任何人都不可能，也绝无必要对它们一一进行跟踪观测。因此，除非意识到或怀疑到自己所观测的有可能不是恒星，天文学家们通常是不会随意对一个天体进行跟踪观测的，而如果不进行跟踪观测，就无法发现行星的运动，从而也就失去了从运动方式上辨别行星的机会。

那么赫歇耳为什么会想到要对天王星进行跟踪观测呢？正是因为他意识到了自己所观测的有可能不是恒星。如我们在第3章中所介绍的，赫歇耳在发现天王星的过程中换用了几种不同的镜片，放大率从227倍增加到932倍⁽¹⁾，从而不仅发现了天王星的圆面，而且还发现其线度随放大率的增加而增加。因此他在静态条件下就发现了天王星与恒星的差别。这是历史上所有与天王星擦肩而过的天文学家们从未有过的优势。以英国的天文学家为例，当时英国皇家天文台最好的望远镜的放大率也只有270倍。赫歇耳拥有如此巨大的设备优势，他成为发现天王星的第一人也就绝非偶然了。而最终使这一伟大发现成为必然的，是赫歇耳所进行的巡天观测。这样的巡天观测正是赫歇耳所说的“品读大自然所写的伟大著作”，在这样周密而系统的“品读”中，一颗像天王星那样的6等星的落网是必然的。



不过，命运有时会跟人开残酷的玩笑。在赫歇耳之前曾经记录过天王星的所有天文学家中，最令人惋惜的是一位法国天文学家，他叫拉莫尼亚（Pierre Charles Le Monnier）。自1750年之后，他先后12次记录了天王星的位置。其中从1768年12月28日到1769年1月23日的短短二十几天里，他不知出于何种考虑，竟然8次记录了天王星的位置。照理说，这样密集的记录是足以显示天王星的行星运动的。但是命运女神却向可怜的拉莫尼亚开了一个最最残酷的玩笑。我们知道，由于地球本身在绕太阳运动，我们在地球上观测到的行星在背景星空中的运动实际上是它们相对于地球的表观运动。对于像天王星这样轨道位于地球公转轨道之外，从而轨道运动速度低于地球轨道运动速度的行星来说，它的表观运动方向有时会与实际的公转方向相反。这就好比当我们坐在一辆正在行驶的车里观测其他车辆时，如果我们自己的车速比较快，就会看到一些与我们同向行驶的车辆相对于我们在倒退。在天文学上，这样的表观运动被称为表观逆行（图4）。表观逆行在行星的表观运动中只占一小部分。在行星从表观逆行转入正向运动的过程中，会有一小段时间看上去是几乎不动的。这就好比一辆倒行的汽车在转为正向行车的过程中，会有一小段时间看上去速度为零。拉莫尼亚万万没有想到的是，他那8次密集记录竟然恰好是在天王星从表观逆行运动转为正向运动的那一小段时间附近，那时候的天王星相对于背景星空几乎恰好是看起来不动的[\[2\]](#)！如果说赫歇耳成为天王星的发现者有什么偶然性的话，这也许就是最大的偶然性。

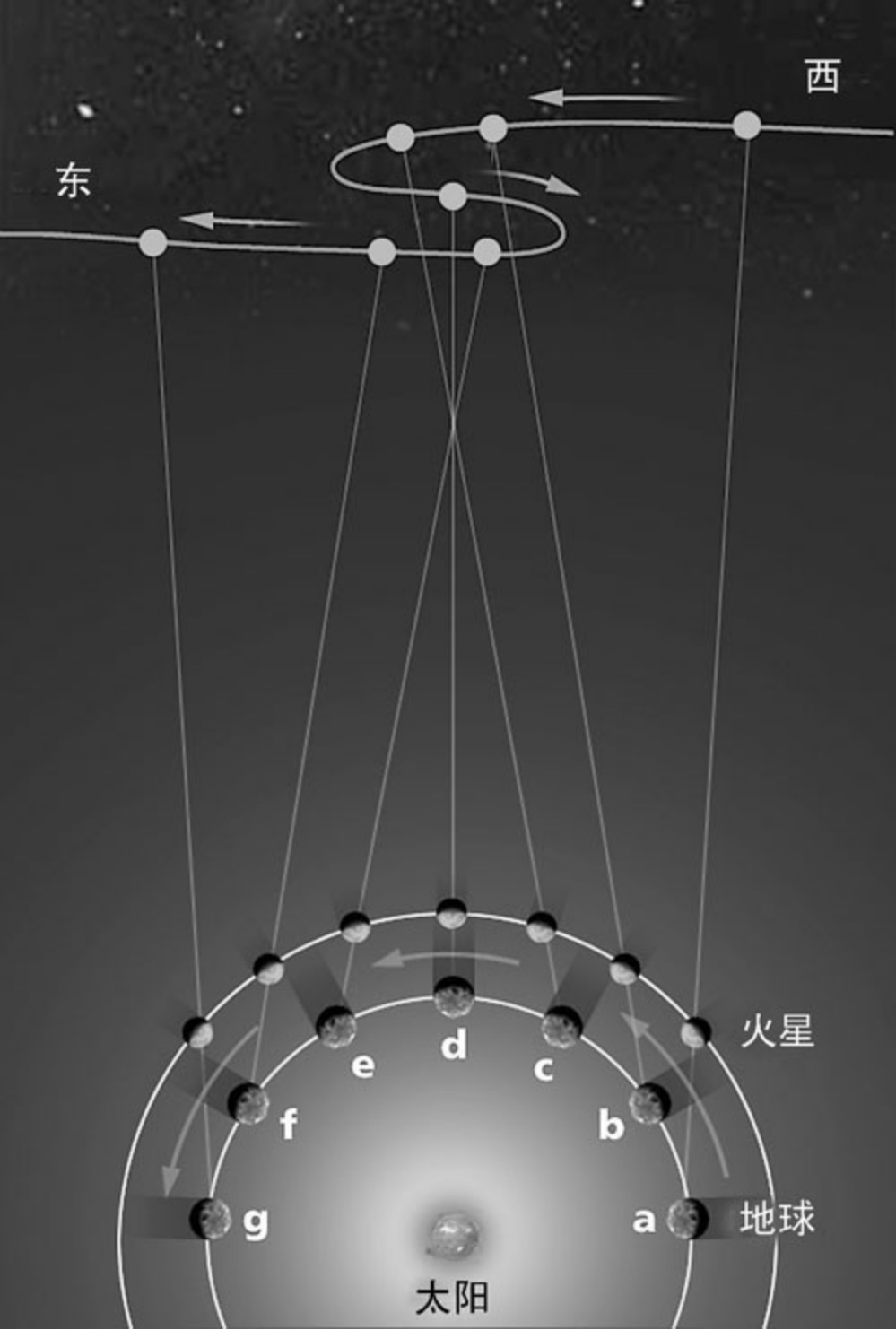


图4 行星的表观逆行

注释

[\[1\]](#) 这还不是赫歇耳当时拥有的最高放大率，后者高达2010倍，比英国皇家天文台最好的望远镜高出将近一个数量级，一度让他的同时代人觉得匪夷所思，有人甚至怀疑那是胡吹。为了平息怀疑，赫歇耳应邀将自己的望远镜带到皇家天文台与那里的望远镜进行了比较。比较的结果是赫歇耳当之无愧地坐上了当时望远镜制作的头把交椅。在比较的过程中最有戏剧性的是马斯克林（即那位最早猜测天王星是行星的天文学家）的反应。他在刚看到“七英尺望远镜”时对它的镜架很感兴趣，打算为自己的望远镜也配备一个，但在比较了两架望远镜的性能后，却沮丧地承认自己的望远镜也许根本就不配拥有一个好的镜架。

[\[2\]](#) 拉莫尼亚的性格比较暴躁，人缘也不好，被普遍视为是一位不细心的观测者，这一点曾被认为是他未能发现天王星的原因。不过有关他“不细心”的某些具体传闻，比如说他将有关天王星的数据随手写在一个纸袋上，实际上是讹传。

5 虚席以待

一颗自17世纪末以来就被反复观测过的6等星竟会是太阳系的第七大行星，赫歇耳的这一发现不仅一举击碎了太阳系行星数目亘古不变的神话，而且激起了人们对寻找太阳系疆界的极大兴趣。“新行星”这一概念几乎在一夜间就从被人遗忘的垃圾股变成了万众瞩目的绩优股，引发了天文学家们极大的热情。在太阳系中，像天王星这样“大隐隐朝市”的行星究竟还有多少？人们恨不得立刻就揭开谜底。

星海茫茫，到哪里去寻找新行星呢？难道要像赫歇耳一样再来一次巡天偶得？幸运的是，太阳系行星的分布就像地球上居民的分布，有一定的规律可循。其中最显著的规律就是行星轨道大都分布在黄道面（即地球的公转轨道平面）附近。这表明，寻找新行星不必漫天撒网，而只需在黄道面附近寻找——这就好比在地球上寻找一位居民时，无需掘地三尺，也不必潜入深海。更幸运的是，行星的分布似乎还有着进一步的规律，这规律帮了天文学家们的大忙。

这个规律的发现可以回溯到天王星发现之前的1766年。那一年，德国天文学家提丢斯（Johann Daniel Titius）注意到：如果以地球公转轨道的半径为单位（这称为天文单位），那么各大行星的轨道半径近似地满足一个非常简单的数学关系式： $r_n = 0.4 + 0.3 \times 2^n$ ，其中：

水星对应于 $n = -\infty$ ， $r_n = 0.4$ （观测值为0.4）；

金星对应于 $n = 0$ ， $r_n = 0.7$ （观测值为0.7）；

地球对应于 $n = 1$ ， $r_n = 1.0$ （观测值为1.0）；

火星对应于 $n = 2$ ， $r_n = 1.6$ （观测值为1.5）；

木星对应于 $n = 4$ ， $r_n = 5.2$ （观测值为5.2）；

土星对应于 $n = 5$ ， $r_n = 10.0$ （观测值为9.5）。



德国天文学家 提丢斯（1729-1796）

这个经验法则除了对火星和土星有5%~7%的偏差外，对其他几个行星都很准确⁽¹⁾。提丢斯将这一结果加注在了自己1766年翻译的瑞士博物学家波涅特（Charles Bonnet）的著作《自然的沉思》中，但在加注时未曾标明自己的名字⁽²⁾。

提丢斯匿名加注的这些结果起初并未引起人们注意。但6年后的1772年，德国天文学家波德，即我们在第3章中提到的那位后来在天王星的命名竞赛中胜出的波德，在为自己的热门著作《星空知识指南》准备新版时，注意到了提丢斯加注在《自然的沉思》中的经验法则。他立

刻被这一法则所吸引，将之添加到了自己的著作中。但很不应该的是，波德在添加这些内容时完全没有提及波涅特或提丢斯的名字。不提提丢斯倒也罢了，因为提丢斯在加注那些内容时是匿名的，可是连波涅特的名字也不提，波德在这件事情上是有显著的剽窃之嫌的。

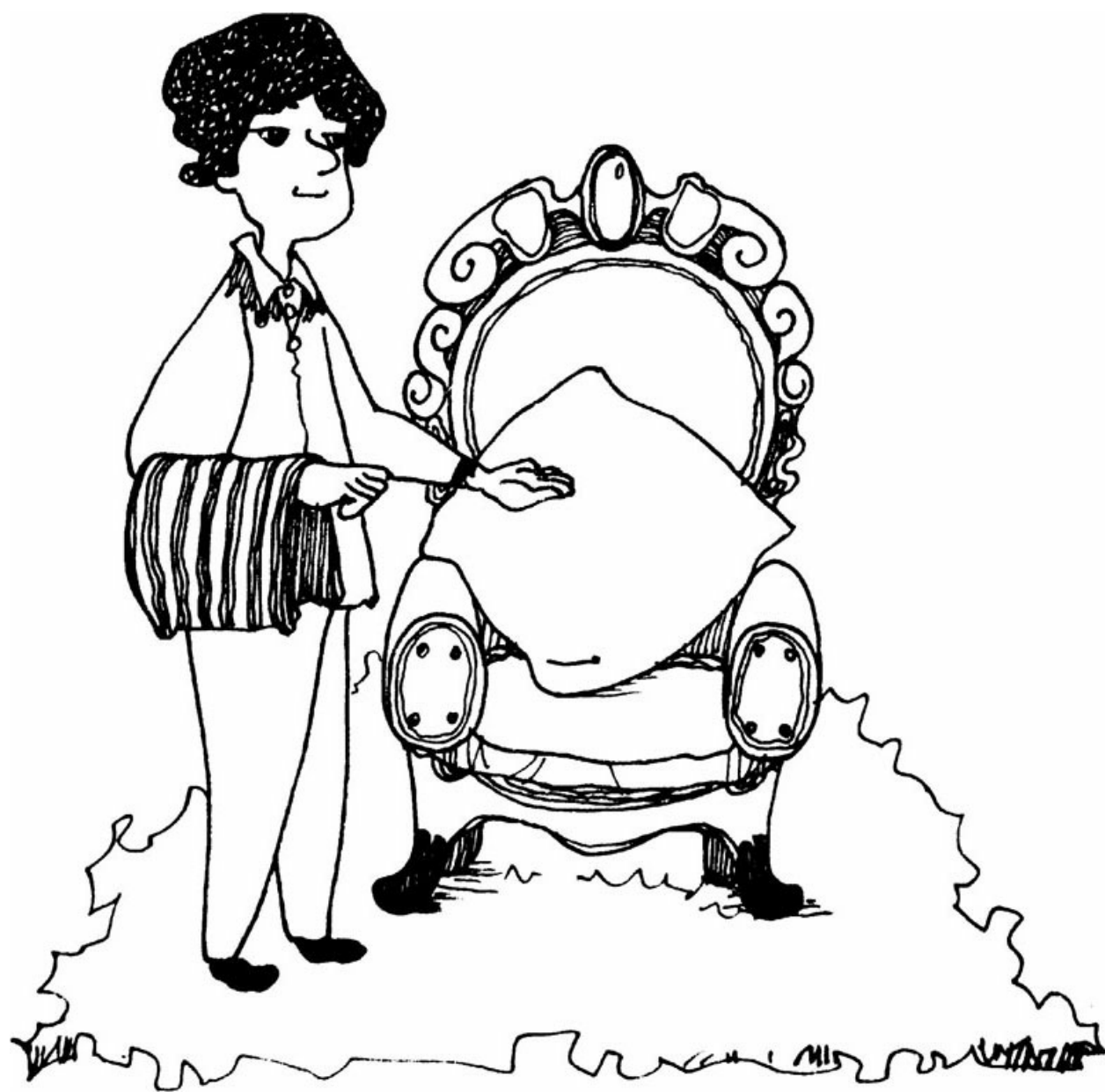
波德的《星空知识指南》在当时受到热烈欢迎，加上波德本人此后几年的积极宣传，在客观上大力传播了提丢斯的经验法则，使波德这位不太光彩的“热心人”成了这一传播的最大受益者，这个经验法则很快就被张冠李戴成了“波德定则”。9年之后，天王星的发现给了波德定则一个极大的支持，天王星的轨道半径与波德定则有着极好的吻合，误差只有2%（请读者自行查验）。这一点使得许多原本认为波德定则纯系巧合的天文学家深受震动，也使波德定则成为后来几十年间天文学家们寻找新行星的重要向导。随着波德定则重要性的提升，历史的真相也开始得到了显现。1784年，在“借用”提丢斯的结果整整12年之后，波德终于承认了提丢斯的贡献。但那时生米早已煮成熟饭，波德的名字与提丢斯的定则已变得难舍难分，后世的天文学家们往往折中地将这一定则称为提丢斯—波德定则。



德国天文学家波德

现在让我们回到寻找新行星的宏伟大业上来。细心的读者或许已经从前面列举的行星轨道数据中看出了一个问题，那就是火星和木星这两个相邻行星的轨道在提丢斯—波德定则中分别对应于 $n=2$ 和 $n=4$ ，中间在距太阳2.8天文单位的地方缺了一个 $n=3$ 。大自然怎么会在火星和木星之间留下如此显著的一个空缺呢？这个问题提丢斯在提出他的定则时就注意到了。这个奇怪的空缺似乎是在虚席以待一颗尚未露面的新行

星，但当时天王星尚未被发现，太阳系六大行星的观念还根深蒂固，提丢斯未敢在太岁头上动土，于是他猜测那里可能存在一颗火星或木星的卫星⁽³⁾。这个猜测很大胆，但也很荒唐，且不说如此远离行星的“卫星”能否稳定地存在，即便真有那样的卫星，又如何能用来填补属于行星轨道的空缺呢？这不成了“指鹿为马”吗？更何况卫星的轨道是以行星为中心的，它与太阳的平均距离与相应的行星与太阳的平均距离相差无几，从数值上讲也根本不可能对应于 $n=3$ 的空缺。波德对这个空缺也很着迷，不过他在这点上比提丢斯略胜一筹，在“借用”提丢斯的结果时，他果断地将提丢斯那破绽百出的卫星猜测改成了行星猜测。



显然，如果提丢斯一波德定则可以信赖，那么寻找新行星的首选战场就应该是火星与木星之间距太阳2.8天文单位的这一神秘空缺。相对于遥远的外行星，这一空缺距离地球可算是近在咫尺，观测起来也相对容易许多。于是天文学家们纷纷将目光汇聚到了那里。

在那里，他们将会发现什么呢？

注释

[\[1\]](#) 当然，这里采用的是现代的表述方式，提丢斯本人的表述是这样的：“将太阳到土星的距离分成100份，那么水星与太阳被4个这样的部分所分隔；金星被 $4+3=7$ 个这样的部分所分隔；地球被 $4+6=10$ ；火星被 $4+12=16$ ；……所分隔。”

[\[2\]](#) 直到1772年再版后，提丢斯才用一个字母“T”（他的姓氏首字母）标明自己所注的内容。而到了1783年，不知是否是出于对波德“借用”其成果的不满，他又过分慷慨地将自己发现的这一经验规律归功于德国哲学家沃夫（Christian von Wolff），其实沃夫只是曾经列出过行星轨道半径的相对大小，并未提出或暗示过任何经验规律。

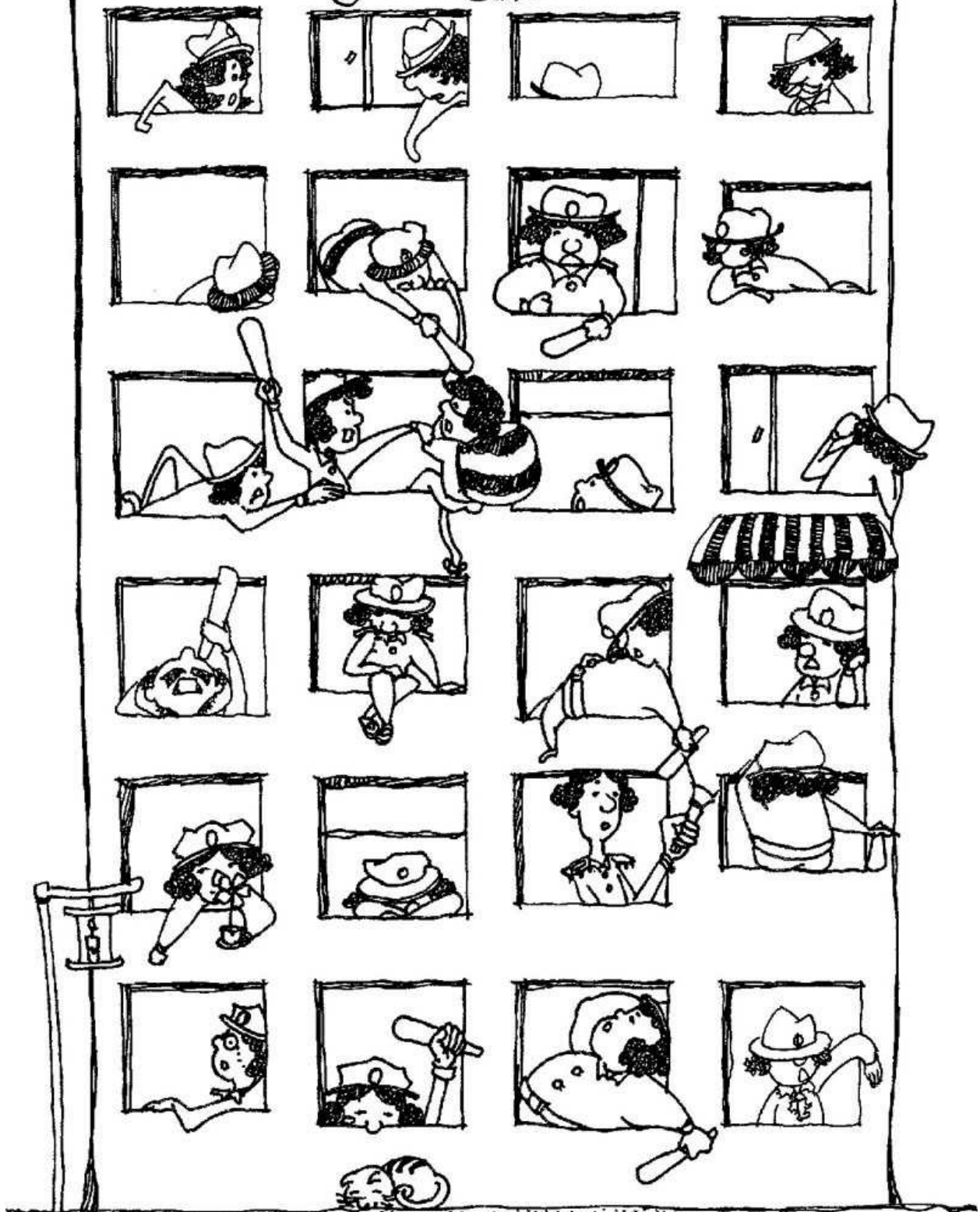
[\[3\]](#) 提丢斯虽然未敢在太岁头上动土，不过比提丢斯更大胆的人也是有的。事实上，早在16世纪末，开普勒就曾猜测过火星与木星之间存在着行星（他还猜测水星与金星之间也存在行星）。在提丢斯之前大约5年，德国哲学家兰伯特（Johann Heinrich Lambert）也曾猜测过火星与木星之间有行星。

6 失而复得

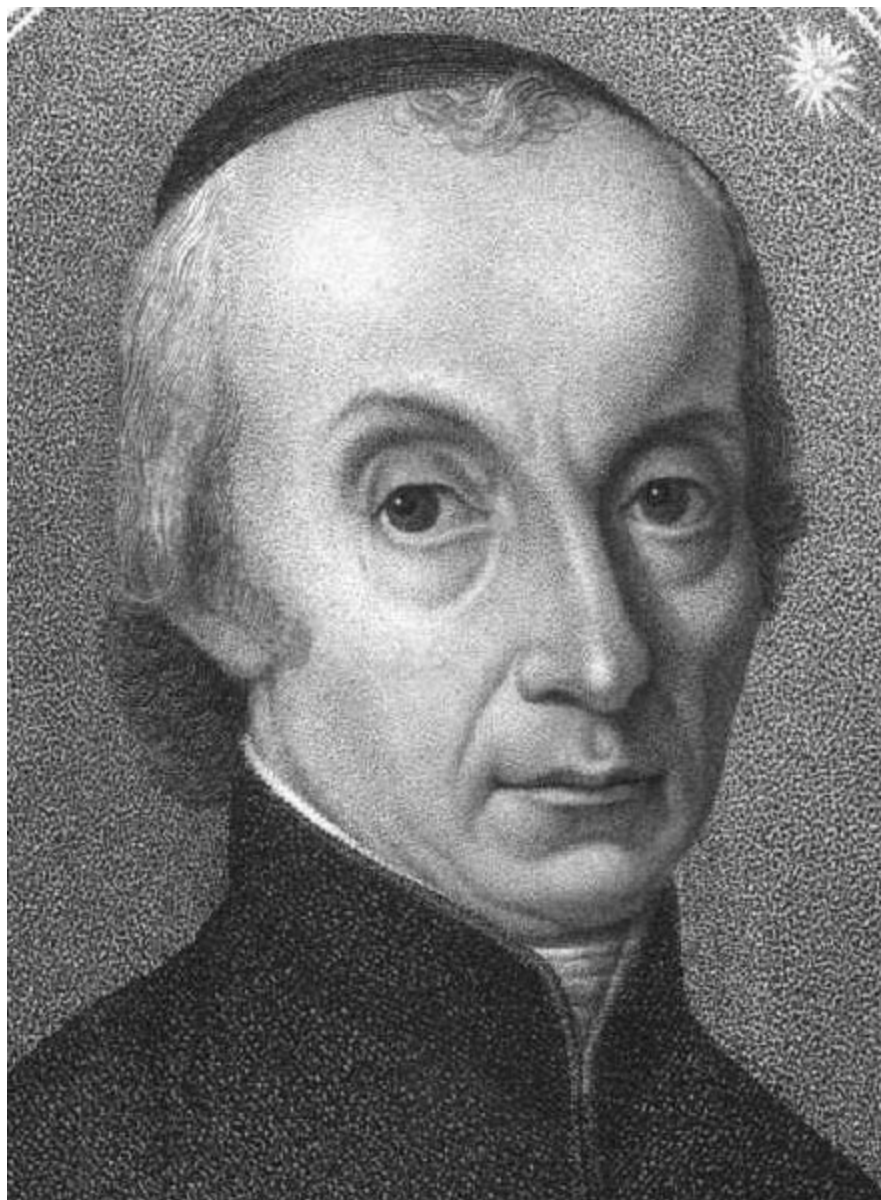
星空的浩渺对于没有真正体验过它的人来说是不容易想象的。即便知道了距离，以及大致的轨道平面，即便离地球如此之近，寻找一颗新行星依然不是一件容易的事情，因为行星出现在轨道的哪一段上仍然是未知的。这就好比警察抓捕逃犯，即便知道逃犯一定就在某座城市里，要想抓住依然不是一件容易的事情，因为逃犯躲在城市的哪个角落仍然是未知的。

当时有意在夜空中抓捕“逃犯”的“警察”还真不少，其中有位叫做扎克（Franz Xaver von Zach）的匈牙利人尤为热心。他曾经拜访过赫歇耳，并从此对寻找新行星产生了浓厚兴趣。自1787年以来，扎克花了整整13年的时间试图寻找位于火星与木星之间的新行星，却一无所获。眼看着“逃犯”将要安然度过18世纪，扎克意识到单枪匹马抓捕“逃犯”的效率实在太低，便决定改变策略。他找了几位运气跟他差不多坏的伙伴商议了一下，决定将新行星的轨道区域分为24块，分别交由24位“天空警察”进行分片搜索。布下这样的天罗地网，无论狡猾的“逃犯”躲在哪个角落都将会难以遁形。老实说，这个分片包干的金点子并非扎克的首创，而是以前就有人提议过，只不过从未付诸实施。

Police



出人意料的是，正当扎克广发英雄帖给欧洲各地的天文学家，抓捕计划已如箭在弦的时候，从意大利的西西里岛（Sicily）忽然传来了“逃犯”落网的消息！勇擒“逃犯”的是一位单枪匹马的“明星警察”，名叫皮亚奇（Giuseppe Piazzi），他当时从事的工作并不是“抓逃犯”，而是“查户口”——为6700多颗星星确定坐标。这是一项枯燥而繁重的工作，为了完成这项工作，皮亚奇一片一片有规律地巡视着星空，在这点上他很像当年的赫歇耳。他这苦力活一干就是11年。1801年1月1日，新世纪来临后的第一天，皮亚奇的望远镜指向了金牛座。这个星座真是天文学家们的幸运星座，20年前赫歇耳就是在这附近发现了天王星，而此刻“户籍警”皮亚奇也在这里迎来了自己一生的一个重要时刻。他对这一小片天区中的50颗星星的坐标进行了记录，第二天，当他对这些星星进行复核时，发现其中有一颗暗淡星体的位置发生了移动！为了确定这种移动不是观测误差，皮亚奇立即对这一天体进行了跟踪观测，结果证实了这种移动的确是天体本身的移动。



意大利天文学家 皮亚奇（1746-1826）

1月24日，皮亚奇写信向同事波德、拉兰德（Joseph Lalande）及挚友奥里安尼（Barnaba Oriani）宣布了自己的发现。为了谨慎起见，他在给波德和拉兰德的信中将自已发现的天体称为彗星。毫无疑问，这是一个与赫歇耳将天王星称为彗星同样的错误。不过在经历了天王星的发现后，皮亚奇比赫歇耳要稍稍大胆一点，他在给挚友奥里安尼的信中指出这个天体有可能是一个“比彗星更好”的东西，因为它的运动缓慢而均

匀，并且不像彗星那样朦胧。为了最终确定这个天体的性质，皮亚奇决定进行更多的观测，并计算它的轨道。可惜他的观测只进行到2月11日就因病中止了。而这时波德、拉兰德及奥里安尼尚未收到他的信件。等那三位收到姗姗来迟的信件，想要确认皮亚奇的观测结果时，新天体已经运动到了太阳附近，消失在了光天化日之中。

虽然失去了当场验证的机会，但波德（他直到3月20日才收到皮亚奇的信）坚信那就是自己期待已久的新行星。当然，相信归相信，最终的判断只能留给观测。好在新天体是不可能一辈子躲在太阳背后的，至多几个月，它必将重返夜空。可问题是：那时候到哪里去找回这颗暗淡的新天体呢？事实证明，这个问题并非杞人忧天，这颗“越狱逃亡”的新天体并没有因为留下过案底就变得容易寻找。日子一天天流淌着，无论天文学家们如何努力，皮亚奇的新天体却再也没有露面。

有读者可能会问：皮亚奇不是对新天体进行了跟踪观测吗？从他的观测数据中把新天体的轨道计算出来不就行了？这个想法是一点都不错的，可实际做起来却绝非易事。在新天体失踪的那些日子里，扎克（他也深信皮亚奇的新天体就是自己想要寻找的新行星）的学生伯克哈特（Johann Karl Burckhardt）就曾对新天体的轨道进行了计算。按照他的计算，天文学家们采取了突击搜查，可惜却扑了个空。皮亚奇自己也进行过计算，结果也劳而无功。计算新天体的轨道之所以困难，是因为皮亚奇的观测只持续了一个多月，所涵盖的只是新天体公转周期的2%左右，而且其中还很不凑巧地包含了表观逆行部分，使结果变得更为复杂。要从这样的观测片断中推算出整个轨道来，无疑是很困难的。更何况观测总是有误差，从这么少的观测数据来推断轨道极易造成误差的放大。最后，我们也不能忘记当时还没有计算机，所有的计算都要依靠纸和笔来完成，这样的计算动辄就要花费很长的时间，有时甚至还不如拿起望远镜直接碰运气来得快捷。因此，推断新天体的轨道，从而预测新天体的位置虽然不是不可能，但却需要福尔摩斯般的技巧，只有第一流

的数学高手才能将这种可能性变为现实。

幸运的是，当时就有一位这样的数学高手前来助人为乐。此人还不是一般的高手，他就是人类有史以来最伟大的数学天才之一，被后人尊称为“数学王子”的德国数学家高斯（Carl Friedrich Gauss）。

当天文学家们为寻找皮亚奇的新天体而忙碌时，这位当时才24岁的数学天才决定助他们一臂之力。在这个节骨眼上，由高斯这样的数学巨匠（虽然当时的高斯还不像后来那么有名）来帮天文学家们计算一个小小的天体轨道，简直就像是摇滚巨星跑来替一家小酒馆义演。高斯仅用两个月的时间，就不仅计算出了新天体的轨道，而且提出了比旧方法高明得多的一整套计算轨道的新方法^{[\(1\)](#)}。高斯把他的计算结果寄给了扎克，后者欣喜若狂，立即公诸于世。借助高斯的计算结果，扎克于1801年12月7日重新找到了皮亚奇的新天体。经过持续观测，他终于在1802年的新年钟声即将敲响的那个夜晚确认了新天体的二度落网，它的位置与高斯的预测只差半度。几个小时之后，德国业余天文学家奥伯斯（Heinrich Wilhelm Olbers）也独立地确认了同样的发现。这颗在1801年的第一个夜晚被“抓获”，又在同一年的最后一个夜晚被重新“捉拿归案”的新天体被称为色列斯（Ceres）（图5）。这是皮亚奇所取的名字，它是罗马神话中的谷物女神，同时也是皮亚奇所在的西西里岛的保护神。在中文中这一天体被称为谷神星^{[\(2\)](#)}。

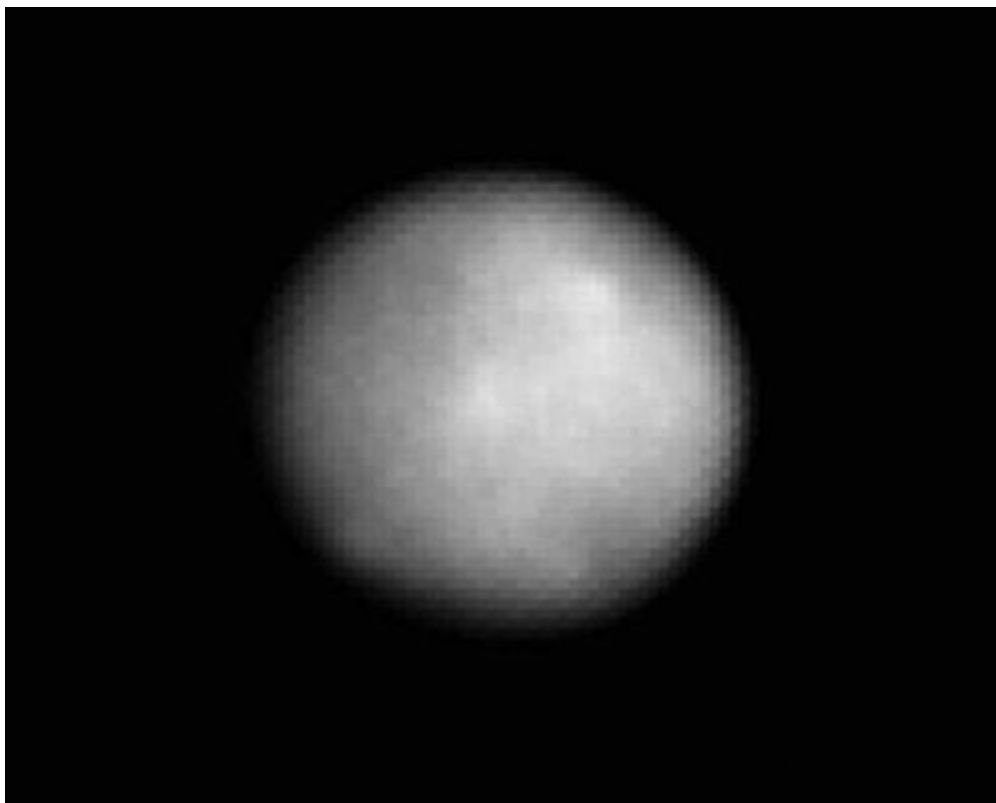


图5 哈勃望远镜拍摄的谷神星

高斯的计算相当精确地给出了谷神星的轨道，它的半径被确定为2.77天文单位，与提丢斯—波德定则吻合得很好（误差只有1%）。看来人们终于找到了位于火星与木星之间的新行星。事实上，早在谷神星被找回之前，对提丢斯—波德定则深信不疑的波德就已急不可耐地将之称为行星了。不过在欣喜之余，天文学家们也感到了一丝困惑：谷神星被皮亚奇发现时的视星等只有8，不仅无法与金、木、水、火、土五大行星相比，甚至比遥远的天王星还暗淡得多。一颗距地球如此之近的行星，为什么会如此暗淡呢？

注释

[\[1\]](#) 高斯在计算中采用了他自己1794-1795年间发展起来的，后来被称为“最小平方”（least square method）的方法。不过他直到1809才发表这一方法，从发表时间上讲，晚于法

国数学家勒让德（Adrien-Marie Legendre），后者1806就发表了最小平方法。

[\[2\]](#) 确切地讲，Ceres只是皮亚奇为谷神星所取名字的前半部分，他提议的全名是Ceres Ferdinandea，其中Ferdinandea是当时那不勒斯和西西里的统治者。与赫歇耳当年提议的“乔治星”一样，Ferdinandea这个带有政治意味的名称也立刻就被天文学家们丢弃了。

7 名分之争

正当人们为谷神星感到困惑的时候，更大的麻烦出现了：1802年3月28日——距离谷神星被重新发现仅仅过了三个多月——与扎克几乎同时找回谷神星的奥伯斯在试图观测谷神星的时候，意外地发现了另外一颗移动的星星。这个天体后来被他称为派勒斯（Pallas），这是希腊神话中的智慧女神，也叫雅典娜（Athena）。在中文中，这一天体被称为智神星。

智神星被发现之后，高斯立刻用自己的新方法计算了它的轨道，结果发现它的轨道虽然不太圆，但平均半径与谷神星几乎完全一样，也是2.77天文单位。这下麻烦大了，一个轨道区域居然挤进了两颗行星，这真是前所未闻的怪事，简直比缺了一颗行星还让人觉得不可思议。如果说发现谷神星带给大家的是喜悦，那么智神星的出现就多少有点令人尴尬了。谷神星的卧榻之侧居然有智神星酣然沉睡，这可能吗？这可以吗？急性子的波德率先对智神星投下了不信任票，他猜测奥伯斯发现的只是一颗彗星（彗星真可怜，总是被人拿来当替代品）。要说波德的这一怀疑还真是有点厚此薄彼，想当初谷神星尚在“越狱潜逃”期间，他就热情万丈地以行星头衔相赠，而现在智神星只不过晚来了几个月，就怀疑人家是彗星世界的“奸细”。

这时候行星观测的元老级人物赫歇耳出来放话了，他说，依我看这两个天体谁也没资格当行星，因为它们都太小了，只能称为小行星（asteroid）[\(1\)](#)。在人们大都期盼新行星的时候，赫歇耳说出这样的话来多少有些扫大家的兴，但以他的身份，说话自然不会是信口开河。那么他的依据何在呢？原来他老人家已经悄悄为这两个暗淡的小家伙度量了“身材”，结果发现它们的直径只有两百多千米。这样的直径还不到月球的1/10，又岂能有资格坐在行星的宝座上？今天我们知道，谷神星的

直径实际有将近1000千米，智神星也有500多千米，远远大于赫歇耳的估计（但仍比月球小得多）（图6）。不过这倒也不能怪赫歇耳，这两个天体实在太小，小到了就连他的望远镜也无法通过观测其圆面来判断大小，而只能通过间接手段进行估计，从而误差很大。不过一开始人们所不知道的是，谷神星的小其实让赫歇耳吃了一个不大不小的哑巴亏：在谷神星“潜逃”的那些日子里，嗜观测如命的赫歇耳也当仁不让地加入到了追捕者的行列，却也像其他人一样铩羽而归。他之所以失败，部分原因就是因为他以为凭借自己天下无双的望远镜，应该能像发现天王星那样直接从圆面上发现谷神星，结果却阴沟里翻了船。谷神星即便在他的望远镜里，也依然保持了苗条的身材，丝毫不显山露水。究其原因，都是太小惹的祸。

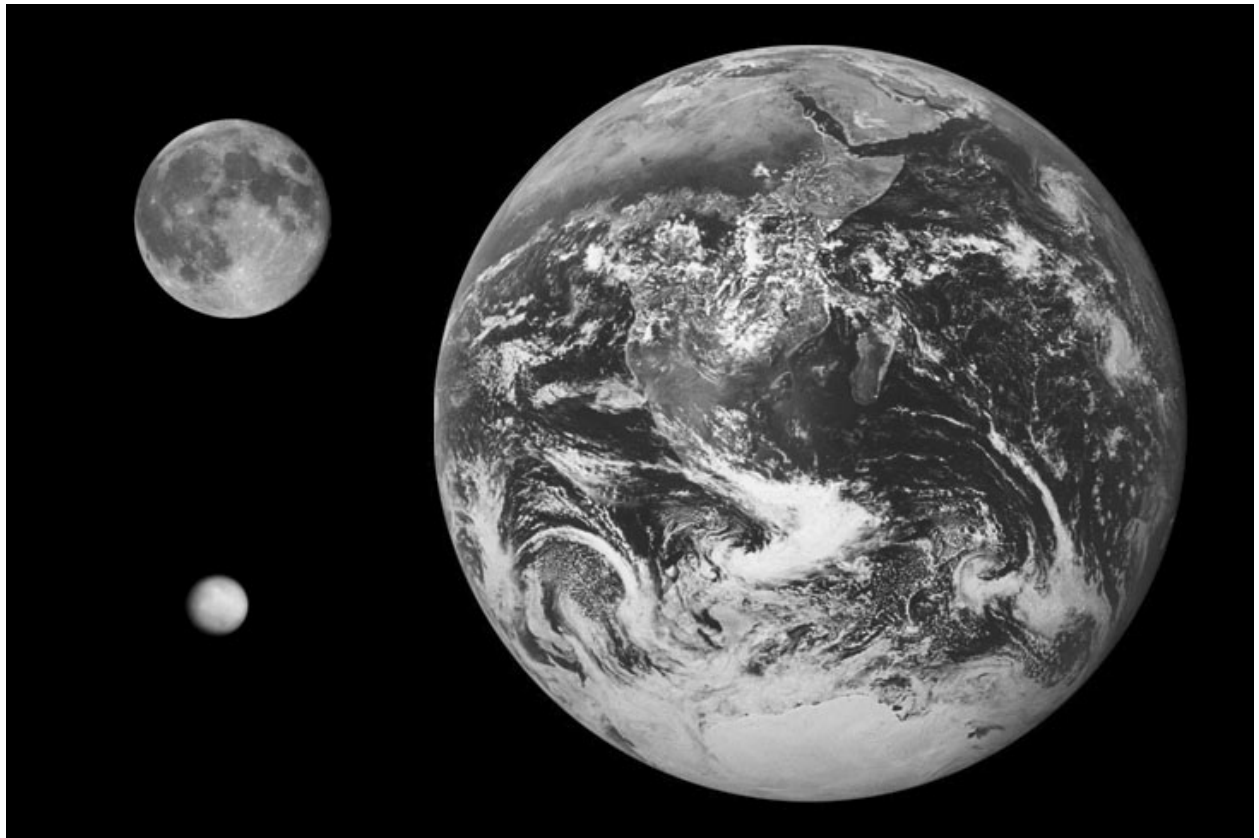


图6 地球（右）、月球（左上）、谷神星（左下）大小对比

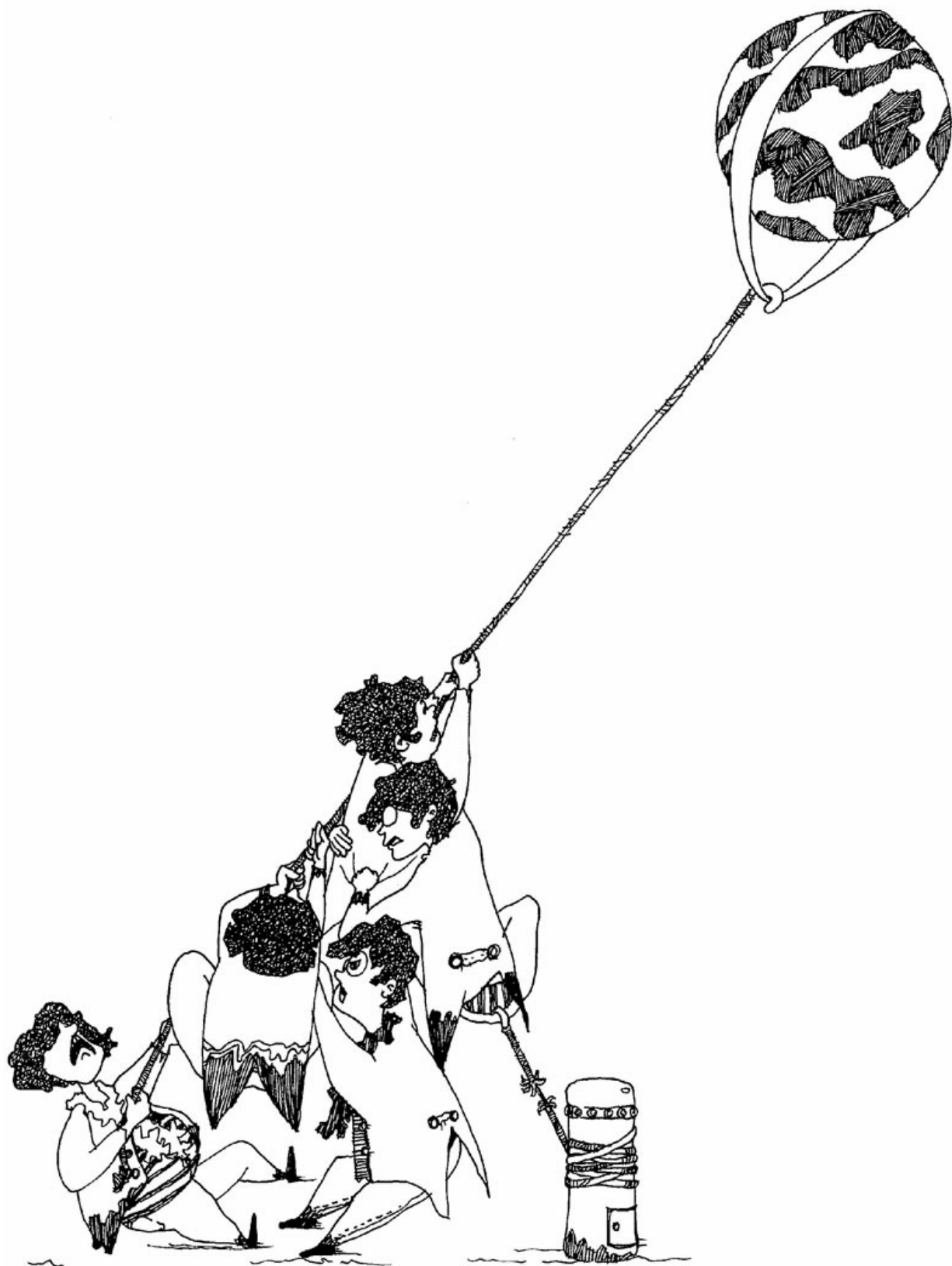


现在该是它为自己的“小”付出代价的时候了。

但赫歇耳的这一提议却遭到了很多人的反对。客气的将之视为文字游戏，不客气的则干脆认为赫歇耳之所以这样提议，目的乃是要让自己发现天王星的贡献盖过皮亚奇和奥伯斯发现谷神星和智神星的贡献（看来荣誉有时还真是一种包袱）。也许归根到底，是人们期待新行星已经期待得太辛苦，实在不想失去已经被发现的新“行星”。不过皮亚奇和奥伯斯这两位发现者本人反倒是没有介意，他们同意了赫歇耳的观点（皮亚奇提议用planetoid取代asteroid，但在谷神星和智神星不具有行星资格这点上他并无异议）。

这场早期的行星名分之争并未持续太久。两年之后，1804年9月1日，德国天文学家哈丁（Karl Harding）在火星与木星之间又发现了一颗新天体，这颗新天体很快被命名为婚神星（Juno），它的轨道也基本满足提丢斯—波德定则的预期。这下算是热闹了，在火星和木星之间抢夺行星宝座的天体由两个变成了三个。不过热闹是热闹了，同时却也成为了最终葬送所有候选者荣登行星宝座机会的导火索。正所谓“三个和尚没水喝”，没有新行星虽然令人失望，可新行星太多了却更让人受不了，于是大家逐渐同意了赫歇耳的提议，将这几个小家伙通通贬为了“小行星”[\[2\]](#)。后来的观测表明，在火星和木星之间存在着成千上万的小行星，它们环绕太阳组成了一个美丽的小行星带。

不过当时谁也不会想到，某些小行星的名分会在时隔两个世纪之后又起了微妙的变化，这是后话。



注释

[\[1\]](#) 赫歇耳提出的名称是“star-like”，意思是“像星星一样”，形容其小。“asteroid”是这一名称在希腊文中的对应。

[\[2\]](#) 这一名分之争的完全落幕其实经历了一个较长的时间。直到1852年，还有天文学教材将当时已发现的小行星与行星合在一起（共计23颗），统称为行星。不过这一趋势在那之后便戛然而止。

8 轨道拉锯

小行星带的发现对提丢斯—波德定则无疑又是一个很大的支持，同时也填补了行星轨道分布中唯一的空缺。如果太阳系还存在其他行星，那么寻找的范围应该是在天王星的轨道之外，对应于 $n=7$ 的地方。这一轨道的半径为38.8天文单位。不过，无论天文学家们对提丢斯—波德定则的信心如何爆棚，一个再明显不过的事实是：即便提丢斯—波德定则真的是一个普遍规律（事实上它并不是），它也绝不可能告诉我们太阳系到底会有几颗行星。提丢斯—波德定则中的 n 可以无限增大，太阳系却不可能是漫无边际的。小行星带由于出现在火星和木星之间的空缺上，因此很多人有理由相信在那里能有所发现。但天王星之外是否存在新的行星，则完全是一个未知数，这使得天文学家们寻找新行星的兴趣在经历了天王星和小行星带的发现之后有所降温。

可惜树欲静而风不止，老天爷看来并不想让天文学家们的日子过得太平静。

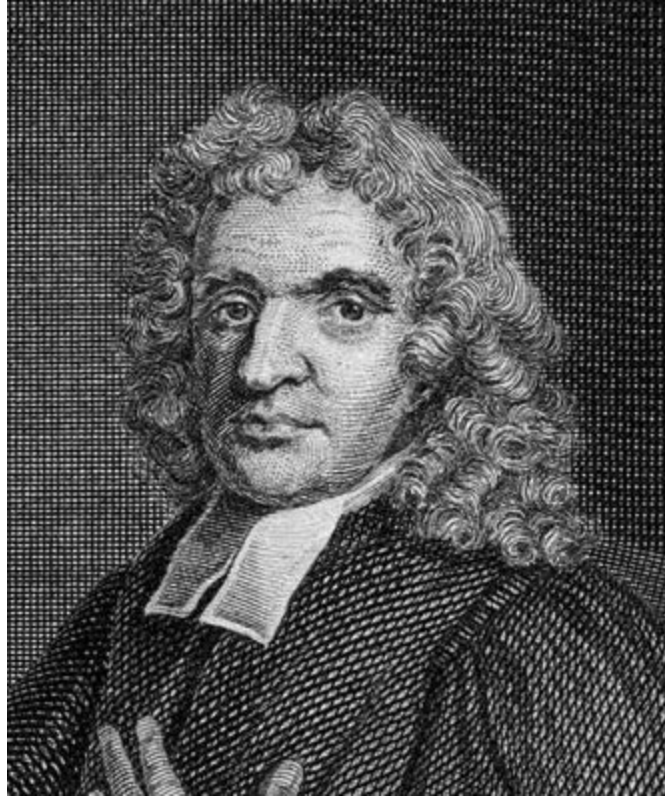
天王星被发现之后，摆在天文学家们面前的一个显而易见的任务就是计算它的轨道。这在当时是很受青睐的工作，这项工作几乎立刻就展开了。如我们在第3章中所说，在短短几个月内，萨隆、莱克塞尔和拉普拉斯就各自计算出了天王星的近似圆轨道，这对于确定天王星的行星地位起了重要作用。两年后，拉普拉斯和他的法国同事梅尚（Pierre Méchain）又率先计算出了天王星的椭圆轨道。

计算出了轨道，人们就可以预言天王星在每个夜晚的位置。一颗遥远行星在天空中的舞步居然可以用科学家手中的纸和笔来导演，这是牛顿力学最令人心醉的地方，也一直是天文学家们在艰苦计算之余最大的欣慰和享受，那种惬意的感觉，宛如是在劳作之后品尝一坛醇香四溢的美酒。不幸的是，这美酒在天王星这里却变了味。当天文学家们放下手

中的纸和笔，将望远镜指向理论预言过的位置，打算像往常一样欣赏一次理论与观测的完美契合时，这位太阳系的新成员却出人意料地缺席了。

天王星的缺席让天文学家们感到了一丝意外。但他们没有想到的是，这小小的意外竟是他们与天王星之间一场长达数十年的拉锯战的开始。

天文学家们起先并未对天王星的缺席太过担忧，因为天王星的轨道周期长达84年，而当时积累的观测数据只有区区两年，还不到轨道周期的3%，凭借这么少的数据是很难进行精确计算的。那么怎样才能改善计算的精度呢？显然需要更多的数据。可积累数据需要时间，这却是半点也着急不得的。怎么办呢？波德想出了一个好主意，那就是翻旧账，看看天王星是否在赫歇耳之前就曾经被天文学家们记录过。如果记录过，那么将那些历史记录与自赫歇耳以来的现代数据合并在一起，就可以既提高计算的精度，又避免漫长的等待。这个一举两得的好主意没有让波德失望，如我们在第4章中所说，天王星的确在赫歇耳之前就曾被反复记录过，其中最早的记录是英国天文学家弗拉姆斯蒂德（John Flamsteed）留下的，时间是1690年，比赫歇耳早了91年。



英国天文学家 弗拉姆斯蒂德（1646-1719）

在历史记录与现代数据的共同帮助下，奥地利天文学家菲克斯米尔纳（Alexander Fixlmillner）率先对天王星轨道作了重新计算，他的计算包含了1690年弗拉姆斯蒂德的记录、1756年梅耶（Tobias Mayer，德国天文学家）的记录，以及1781年至1783年间赫歇耳和他自己的观测数据。他的计算与观测数据之间的误差只有几〔角〕秒⁽¹⁾，这在当时是很不错的结果。1786年，菲克斯米尔纳发表了他的结果。在天文学家们与天王星的轨道拉锯战中，菲克斯米尔纳为天文学家们拔得了头筹。

可惜好景不长，菲克斯米尔纳的计算发表后才过了两年，天王星就扳回了一城——它偏离了菲克斯米尔纳的轨道。心有不甘的菲克斯米尔纳尽了最大的努力试图挽救自己的计算，却沮丧地发现历史记录与最新的观测数据仿佛变成了一付跷跷板的两个端点，一端压下去，另一端就会跷起来。看来鱼和熊掌已无法兼得，菲克斯米尔纳决定舍鱼而取熊

掌，他做了一个在他看来最合理的选择，那就是抛弃年代最为久远的弗拉姆斯蒂德的观测记录。做出了这种“壮士断腕”的行动后，菲克斯米尔纳再次计算了天王星的轨道，总算重新将误差控制在了10〔角〕秒以内。

但人们对菲克斯米尔纳的选择并不满意，因为被他抛弃的弗拉姆斯蒂德的记录虽然年代久远，观测手段相对简陋，但信誉却丝毫不容低估。弗拉姆斯蒂德是格林威治天文台的奠基者，也是英国第一位皇家天文学家，不仅拥有显赫的头衔，而且素以观测细心著称。他当年曾为牛顿的巨著《自然哲学的数学原理》提供过大量的观测数据⁽²⁾，他所绘制的星图不仅在当时无与伦比，甚至在一个世纪之后仍被奉为经典。在赫歇耳进行天文观测时，放在他桌上作为参考的正是弗拉姆斯蒂德的星图。因此抛弃弗拉姆斯蒂德的记录于情于理都很不妥当，菲克斯米尔纳的新计算能否算是胜利，实在很难论断。

如果不抛弃弗拉姆斯蒂德的记录，跷跷板却又摆不平，这该如何是好呢？简单的逻辑告诉我们，在观测与理论出现矛盾时，如果观测没有问题，那问题就应该出在理论上。当时的理论确实有一个致命的弱点，那就是只考虑了太阳的引力，而没有考虑其他行星的影响，这其中尤以木星和土星的影响最不容忽视。1791年，法国天文学家达兰伯利（Jean Baptiste Joseph Delambre）率先考虑了这两颗巨行星对天王星轨道的影响。他的计算很好地拟合了当时已知的所有观测数据，其中包括被菲克斯米尔纳抛弃过的弗拉姆斯蒂德的数据，以及不久前才被发现的拉莫尼亚的早期观测数据。



法国天文学家 达兰伯利（1749-1822）

在木星和土星这两位老大哥的坐镇之下，天王星的气焰终于被打压了下去，天文学家们重新夺回了阵地，太阳系也重新恢复了往日的循规蹈矩^{[\(3\)](#)}。这一“和谐太阳系”维持了较长的时间，直到1798年英国天文学家霍恩斯比（Thomas Hornsby）视察战场时，胜利的果实还在枝头挂

着。可就在人们以为战争已然落幕，刀枪可以入库的时候，天王星这个注定不肯让天文学家们平静过完18世纪的家伙，却将枪口重新探出了大幕！



法国天文学家 波瓦德（1767-1843）

自1800年（18世纪的最后一年）起，天王星的轨道开始系统性地偏离达兰伯利的计算。

沉默了8年的天王星不鸣则已，一鸣惊人，而天文学家们的手中却已无牌可打，只得仓皇退避。

这一退堪称惨败，整整20年没缓过劲来。直到1820年，才有一位叫做波瓦德（Alexis Bouvard）的法国天文学家站出来绝地反击。这20年里天文学家们倒也没闲着，现代数据增加了20年自不用说，手头的历史记录也增添了两项：一项是新发现的弗拉姆斯蒂德在1712-1715年间的观测记录，这项记录很好地填补了弗拉姆斯蒂德1690年的记录与拉莫尼亚1750年的记录之间原本长达60年的数据空白；另一项则是英国天文学家布莱德利（James Bradley）1753年的观测记录。这时天文学家手中的数据早已不再匮乏，不仅不匮乏，反而多到了能噎死人的程度。波瓦德稍加检视，就发现自己面临的局面与30年前菲克斯米尔纳曾经面临过的有着惊人的相似：那就是历史记录与现代数据无论如何也不能匹配。30年前的局面还有木星和土星来解围，30年后的今天还能依靠什么呢？无奈之下，波瓦德只得效仿菲克斯米尔纳的“壮士断腕”。可如今的局面比30年前还要糟糕，连断腕都不够，得断臂——将赫歇耳之前的所有历史记录一笔勾销——才行。就这样，波瓦德靠着“壮士断臂”的悲壮，于1821年计算出了一个新轨道，这个轨道与自赫歇耳以来的新数据勉强吻合。

这样的反击能算是成功吗？恐怕连惨胜都算不上吧？人们还从未在一颗行星的轨道计算上栽过如此多、如此大的跟斗。而且这次付出的代价也实在太大了，居然把凝聚了那么多天文学家心血的所有历史记录都丢弃了。即便如此，波瓦德的轨道与某些现代数据的偏差也仍然高达10〔角〕秒左右，这虽不致命，却也令人疑惑。不过对于自赫歇耳以来的新数据而言，这一轨道毕竟是当时最好的，并且事实上也是唯一一个尚堪使用的轨道，聊胜于无，因此一些天文学家还是勉强接受了它。

光有天文学家的接受是没有用的，关键还得看天王星这位“敌人”是否赏脸。

那么“敌人”的回答是什么呢？



注释

[\[1\]](#) [角] 秒是非常小的角度，等于1度的1/3600，或者圆周（360度）的1/1296000。

[\[2\]](#) 弗拉姆斯蒂德后来与牛顿闹翻了，此后牛顿利用其至高无上的地位，以种种不甚光彩的手段对弗拉姆斯蒂德进行了打击。

[\[3\]](#) 在这几年中，天文学家们在人间的日子却过得很不平静，在法国大革命最血腥狂热的1794年，最早计算出天王星圆轨道的萨隆死于断头台。

9 众说纷纭

在这个节骨眼上“敌人”倒是很沉得住气，没有立刻表态。但仗打到这个份上，“敌人”的不置可否反倒让天文学家们无所适从，几乎陷入了“窝里斗”。事实上，很多天文学家对波瓦德付出的“断臂”代价耿耿于怀，因为按照波瓦德的轨道，那些被丢弃的历史记录与计算之间的偏差高达几十〔角〕秒⁽¹⁾。这么大的偏差居然同时出现在这么多彼此独立的观测结果中，难不成留下历史记录的那些天文学家全都在观测天王星的时候喝了酒？这实在是令人难以置信的事情。就连波瓦德本人也不得不承认，造成历史记录与现代数据无法匹配的真正原因有待于后人去发现。

但耿耿于怀也好，难以置信也罢，“敌人”既然没有反对，天文学家们也不便自己拆自己的台，于是有人开始为波瓦德抛弃历史记录的做法寻找可能的解释。其中有一种解释认为天王星曾经被某颗彗星“撞了一下腰”，从而偏离了正常的轨道⁽²⁾。如果是这样，那么历史记录与现代数据无法匹配就不再是问题了，因为它们描述的分别是碰撞前和碰撞后的轨道，本来就应该彼此不同。由于历史记录一直覆盖到1771年（那是拉莫尼亚的最后一次记录），而现代数据则开始于1781年（那是赫歇耳的第一次观测），因此人们猜测该撞击发生在1771-1781年间。

但是像彗星撞击这样为了特定目的而提出的建立在偶然事件基础上的假设，是科学家们素来不喜欢的。因为人们若是时常用这类假设来解释问题的话，科学就会变成一堆零散假设的杂乱集合，而丧失其系统性。更何况彗星撞击天王星不仅概率实在太小，而且由于彗星的质量与天王星相比简直就是九牛一毛（事实上比九牛一毛还远远不如），即便真的撞上天王星，也万万不可能对后者的轨道产生任何可以察觉的变化。反过来说，倘若真有一个天体可以通过撞击天王星而显著改变其轨

道，那么该天体的质量必定极其可观，那样的撞击若果真发生在1771-1781年间，绝对会是一个令人瞩目的天象奇观，又怎可能不留下任何直接的观测记录呢？因此，彗星撞击说从各方面讲都是一个很糟糕的假设。连这样糟糕的假设都被提了出来，天文学家们在天王星问题上的处境之绝望可见一斑。

更糟糕的是，即便在那样的处境下，天王星还是毫不手软地往天文学家们的伤口上撒了一把盐。自1825年起，天王星故伎重演，开始越来越明显地偏离波瓦德的轨道。几年之后，两者的偏差已经达到了令人绝对无法忍受的30〔角〕秒。

“敌人”那姗姗来迟的回答终于被等到了，可惜却是一个那么残酷的回答。

这时候天文学家们实已一败涂地，而且还败得极其难看，因为天王星早不出手晚不出手，偏偏是在天文学家们“臂”也断了，“血”也流了，还煞费苦心地为自己的断臂找了借口之后才出手。那情形，怎一个“惨”字得了？当然，到了这时候，人们也已经习惯了，失败早已不是新闻，天王星若是乖乖听话了反倒会成为新闻。

屡战屡败之下，天文学家们开始改换思路。

仔细想想，彗星撞击说虽然很失败，却也并非一无是处，起码在思路，它尝试了用外力的介入来解释天王星的出轨之谜。沿着这样的思路，天文学家们又提出了另外一些假设，比如认为天王星的出轨是由某种星际介质的阻尼作用造成的。这种假设以前曾被用来解释某些彗星的轨道变化，但用它来解释天王星的出轨却面临一个致命的困难，那就是介质的阻尼作用只能阻碍天王星的运动，而绝不可能起到相反的作用。说白了，就是只能让天王星运动得更慢，而绝不可能相反。但不幸的是，天王星的运动却有时比理论计算的慢，有时却比理论计算的快，这样的偏差显然是不可能用介质的阻尼作用来解释的。

还有一种假设则认为天王星出轨是由一颗未知卫星的引力干扰造成

的。这种假设也有一个致命的弱点，那就是如果真的存在那样的卫星，它的质量应该远比当时已知的两颗天王星卫星大得多，那么大的卫星为何一直未被发现呢？这是很难说得通的。更何况卫星绕行星运动的周期相对于行星的公转周期来说一般都很短（比如月球绕地球运动的周期只有地球公转周期的1/12，对外行星来说这一比例通常更小），由此造成的行星轨道变化应该是短周期的，可是天王星出轨的方式却呈现长期的变化，这是卫星假设无法解释的，因此卫星假设也很快就脸朝下地躺倒了。

除这些假设外，有些天文学家还提出了另外一种可能性，那就是万有引力定律也许并不是严格的平方反比律，甚至有可能与物质的组成有关。这种可能性虽然很难被排除，但万有引力定律是一个“牵一发动全身”的东西，一旦被修正，所有天体的运动都将受到影响，其中也包括那些一直以来被解释得非常漂亮的其他行星及卫星的运动。要想对万有引力定律动手脚，让它解释天王星的出轨，同时却又不破坏其他行星的运动，无疑是极其困难的——如果不是完全不可能的话。而且单凭天王星的出轨就在天体力学祖师爷牛顿的万有引力定律头上动土，也似乎太小题大做了一点，因此这种假设的支持者寥寥无几。

就这样，从拉普拉斯、梅尚、菲克斯米尔纳、达兰伯利，到波瓦德，一次次的计算全都归于了失败；从彗星撞击说、介质阻尼说、未知卫星说，到引力修正说，一个个的假设全都陷入了困境。天王星出轨之谜的正解究竟在哪里呢？到了19世纪30年代末，天文学家们在盘点自己的“假设仓库”时发现那里只剩下了一张牌。这张牌是唯一一个没有倒下的假设，这个假设已是最后的希望，可这个希望的背后却是一道令人望而生畏的数学难题。

天王星出轨之谜的正解期待数学高手的横空出世！

注释

[\[1\]](#) 多数读者可能对天文观测的精度没有概念。几十〔角〕秒的误差是个什么概念呢？那相当于丹麦天文学家第谷（Tycho Brahe）的观测误差。第谷是16世纪的天文学家，比最早观测天王星的弗拉姆斯蒂德还早了一个世纪，他的全部观测都是依靠肉眼进行的（望远镜的发明是他去世之后的事）。因此几十〔角〕秒的误差所对应的精度是肉眼观测的精度（虽然对于肉眼来说这应该算是最高的精度，因为第谷是他那个时代最杰出的天文观测者）。望远镜发明之后，天文观测的精度有了数量级的提高。据分析，伽利略的观测精度就已达到了两〔角〕秒。

[\[2\]](#) 有的读者可能会觉得奇怪，天文学家们怎么每次碰到问题时，都会拿彗星说事？发现新行星时先说是彗星，不想让某个天体（比如智神星）成为行星时也说是彗星，现在又说天王星被彗星撞了腰。原因其实很简单，因为在那时候，天空中最常被观测到的不速之客就是彗星。

10 数学难题

这个硕果仅存的假设读者们想必已猜出来了，那就是在天王星轨道的外面还存在另一颗大行星，正是它的引力作用干扰了天王星的轨道，使它与天文学家们玩了将近半个世纪的捉迷藏。（请读者们定性地想一想，天王星之外存在新行星的假设为何能避免前面提到的介质阻尼说和未知卫星说所遭遇的困难？）在天王星之外存在新行星的猜测本身其实并不出奇。事实上，自天王星被发现之后，稍有想象力的人都可以很容易地想到这一点。不过，泛泛猜测一颗新行星的存在是一回事，将这种猜测与已知天体的运动联系起来，从而形成一种具有推理价值的假设，乃至用这一假设来解决一个定量问题，则是另一回事。后者无疑要高明得多，困难得多，它的出现也因此要晚得多——直到1835年才正式出现。

1835年的11月，著名的哈雷彗星经过了将近76年的长途跋涉，重新回到了近日点。天文爱好者和天文学家们共同迎来了一次盛况空前的天文观测热潮。就在万众争睹这个多数人一生只有一次机会能用肉眼看到的美丽彗星时，天文学家们却注意到了一个小小的细节：那就是哈雷彗星回到近日点的时间比预期的晚了一天。一个长达76年的漫长约会只晚了区区一天，算得上是极度守时了，但天文学家们的敏锐目光并未放过这个细小的偏差。法国天文学家瓦尔兹（Benjamin Valz）和德国天文学家尼古拉（Friedrich Bernhard Nicolai）几乎同时提出了一个假设，那就是哈雷彗星的晚点有可能是受一颗位于天王星轨道之外的新行星的引力干扰所致^{[\(1\)](#)}。由于当时天王星出轨之谜早已传得沸沸扬扬，瓦尔兹进一步猜测这颗未知行星有可能也是致使天王星出轨的肇事者，这便是天王星出轨之谜的新行星假设。



与那些一出道就遭遇致命困难的其他假设相比，新行星假设没有显著的缺陷，这个难能可贵的特点使它很快就脱颖而出。到了1837年，就连波瓦德也开始接受这一假设了。波瓦德的外甥在给英国皇家天文学家艾里（George Biddell Airy）的一封信中提到，他舅舅（即波瓦德）已开始相信天王星出轨的真正原因在于天王星之外的未知行星的干扰

[\[2\]](#)。艾里当时是英国格林威治天文台的台长，他在我们后面的故事中将是一位重要人物。艾里对天王星出轨之谜也非常关注，他手头掌握着大量的观测数据，并且还撰写过有关这一问题的详尽报告。通过对天王星轨道数据的细致分析，艾里发现了一个当时鲜为人知的问题，那就是计算所得的天王星位置不仅在角度上与观测数据存在着众所周知的偏差，而且在径向——即天王星与太阳的距离——上也与观测数据存在偏差。艾里认为这种偏差的存在表明理论计算本身还有缺陷，他把这看成是解决天王星出轨之谜的关键。至于新行星假设，艾里则很不以为然。



英国天文学家 艾里（1801-1892）

艾里对新行星假设的不以为然，以及他对天王星出轨症结的判断后来被证实是不正确的。他个人所持的这些观点虽不足以阻挡新行星假设快速流行的步伐，但由于他在英国天文学界举足轻重的地位，他的这种日益孤立的见解为后来英国在寻找新行星的竞争中落败埋下了种子。

新行星假设虽然受到了广泛关注，但要想证实这一假设却绝非易事。证实它的最直接的方法显然就是找到这颗未知的新行星，通过观测确定其轨道，然后再根据其轨道计算它对天王星的影响。如果这种影响恰好能够解释天王星的出轨，那么新行星假设就算得到了证实。

可问题是，究竟该到哪里去寻找这颗未知的新行星呢？它离太阳的距离比天王星还要遥远（如果提丢斯—波德定则有效的话，它离太阳的距离应该比天王星远一倍左右），因此一般预期其亮度要比天王星暗淡得多。另一方面，它的移动速度则要比天王星慢得多，因此不仅搜寻的难度大得多，而且判断其为行星也要困难得多。这样的搜寻听起来意义非凡，实际上却是一项风险很大的工作，很可能投入了巨大的人力、物力及时间，结果却换得竹篮子打水一场空。另一方面，当时各大天文台都有相当繁重的观测任务（其中有些虽号称是天文观测，其实却是“为国民经济保驾护航”一类的测绘及定位任务），既没有意愿也没有条件进行这种额外并且高风险的搜寻工作。

既然依靠观测这条路走不通，那么对新行星假设的判定就只能通过纯粹的数学计算来实现了。毫无疑问，这种执果求因的计算要比计算一颗已知行星对天王星的影响困难得多，因为新行星既然是未知的，它的质量、轨道半径、轨道形状、与天王星的相对角度等所有参数也就都是未知的。因此在计算时既要通过天王星出轨的方式来反推那些参数的数值（这是相当困难的数学问题），也需要对无法有效反推的参数数值进行尽可能合理的猜测，然后还得依据这些反推或猜测所得到的参数来计算新行星对天王星的影响，并通过计算结果与观测数据的对比来修正参数（这是相当繁重的数值计算，别忘了那时还没有计算机）。这种反推、猜测、计算、对比及修正的过程往往要反复进行很多次，才有可能得到比较可靠的结果。因此要求计算者既有丰富的天体力学知识，又有高超的计算能力，而且还要有过人的毅力、耐心和细致。

幸运的是，历史不仅给了天文学界这样的人物，而且很慷慨地一给就是两位。

注释

[\[1\]](#) 早在1758年，即天王星尚未被发现的时候，就有天文学家猜测像哈雷彗星这样有机会

远离太阳的彗星，有可能受到遥远的未知行星的影响。不过那种猜测在当时并未得到任何具体数据的支持。

[\[2\]](#) 据艾里后来回忆，他甚至在1834年就曾收到过一位名叫赫西（Thomas Hussey）的英国业余天文学家的类似提议，不过那个提议没有明说未知天体是一颗行星。

11 星探出击

就在波瓦德向天王星轨道问题发起唐吉珂德式冲击的前一年，即1819年，一个小男孩降生在了英国康沃尔郡（Conwall）的一个农夫家庭，他被取名为约翰·亚当斯（John Couch Adams）。这个孩子很早就显露出超乎常人的数学计算能力。还在孩提时期，他就自学掌握了大量数学技巧。在16岁那年，他通过复杂的计算相当准确地预言了发生在当地的一次日食，震动了乡邻，也预示着他一生的追求。



1839年，亚当斯进入剑桥大学的圣约翰学院深造。两年后的一个夏日，他在一家书店里看到了艾里有关天王星问题的报告。那时候，观测数据与波瓦德轨道的偏差已达到了创纪录的70〔角〕秒，天王星出轨之谜比以往任何时候都更尖锐。已有18年历史的波瓦德轨道虽已千疮百孔，却仍像幽灵一般浮现在天文学家们的眼前，刺痛着他们。但这一切对于年轻的亚当斯却是一个巨大的机会。对一位16岁就能预言日食的年轻数学高手来说，有什么能比与天王星出轨之谜这样的绝世难题同处一个时代更令人兴奋呢？亚当斯立即就被这一问题深深地吸引住了。

不过，吸引归吸引，年轻的亚当斯还不能马上就投入到这个问题中去。为了替自己今后从事真正的学术研究创造尽可能有利的条件，他必须首先完成剑桥大学的学业，为两年后将要到来的毕业考试做好准备。这些虽不是他的终极兴趣，却对他最长远的学术前途有着至关重要的影响。现实人生往往就是如此，你想做一件事，生活却用这样或那样的事情来牵制你的兴趣。不过处置得宜的话，这种牵制未必会成为羁绊。亚当斯的努力没有白费，1843年，他以最优异的成绩通过了毕业考试，据说他的数学成绩竟比第二名高出一倍以上。几星期后，他又获得了剑桥大学的最高数学奖——史密斯奖，并如愿以偿地成为了圣约翰学院的研究员（fellow）。

1843年的最后几个月，亚当斯的生活甚至比考试前还要繁忙。放在他面前的是三重任务：一是对天王星轨道进行研究，这是他的梦想和兴趣，他终于可以追逐自己的梦想了，但在时间上却必须与其他任务共享；二是要完成圣约翰学院的教学任务，这是生存的需要；三是替剑桥天文台的台长查利斯（James Challis）计算一颗彗星的轨道，这是他与剑桥学术界的正式交流，同时也是一次很好的热身，因为这一计算要求考虑木星对彗星的引力干扰，而天王星出轨之谜的关键也正在于其他行

星的引力干扰，两者不无相似之处（当然后者要困难得多）。亚当斯有关彗星的计算发表于1844年初，他的结果与一位法国天文学家的计算吻合得很好，这一点很让他高兴。但他也许做梦也不会想到，自己与那位法国天文学家的命运在未来几年里竟会交织出那么多的故事和风波。

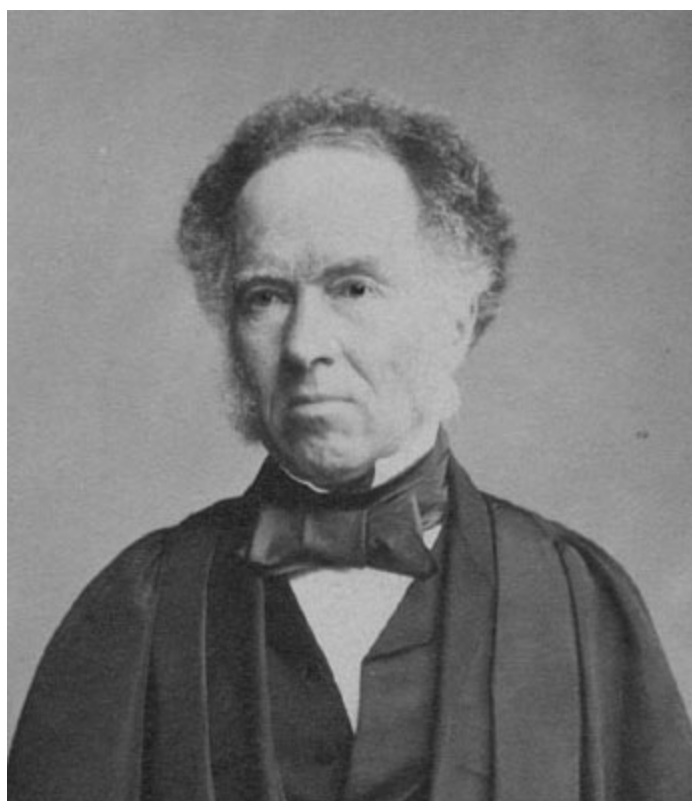


法国天文学家 勒维耶（1811-1877）

那位法国天文学家的名字叫做勒维耶（Urbain Le Verrier），他出生在法国的诺曼底，比亚当斯大8岁。他正是历史带给天文学界的另一位数学高手！

与完成彗星轨道的计算几乎同时，亚当斯也完成了对天王星轨道的

初步分析。他首先仔细检查了波瓦德的计算，发现并纠正了一些错误，但这些小打小闹并不足以挽救波瓦德的轨道。在确信波瓦德轨道已经无可救药之后，亚当斯正式采纳了新行星假设。那颗神秘的新行星究竟在哪里呢？亚当斯开始了用纸和笔寻找答案的艰难历程。作为计算的起点，他假定新行星在提丢斯—波德定则所预言的距太阳38.8天文单位的椭圆轨道上运动。他的初步评估得到了令人振奋的结果：新行星对天王星的影响看来的确可以解释天王星的出轨之谜。但为了得到可靠的结果，亚当斯需要更多的数据，于是他向自己刚刚帮助计算过彗星轨道的查利斯求援。



英国天文学家 查利斯（1803-1882）

查利斯在我们后面的故事中也是一位重要人物，他当时很够意思，立即就写信向艾里索要数据。艾里我们已在上章中提到过，他是当时格林威治天文台的台长，在英国天文学界算得上是重量级的人物。之前，

他也曾担任过剑桥天文台的台长，因而是查利斯的前任。如我们在上章中所说，艾里很关心天王星出轨之谜，手头也有最新的观测数据，但他对新行星假设并不看好。艾里工作一丝不苟，但为人古板，缺乏想象力。在他管束下的格林威治天文台台规森严、条框众多。这一切对后来故事的发展有着不可忽视的影响。亚当斯通过查利斯向艾里索要数据，也许是他与艾里之间第一次打交道。这次交道虽然间接，但却非常顺利，艾里立刻就寄来了数据。可惜这也是接下来两年半的关键时间里亚当斯与艾里之间唯一一次顺利的交道。

拿到了数据，亚当斯立刻就投入到了更精密的计算之中。不过，学院的教学任务与彗星计算还是从他那里夺走了一部分时间。1844年秋天，查利斯又让亚当斯帮他计算一颗彗星的轨道——那是一颗新发现的彗星。那时亚当斯对此类计算早已轻车熟路，秋叶尚未落尽，他的计算结果就出来了。但出人意料的是，亚当斯如此麻利的计算竟然还是慢了一步，一个已不再陌生的法国名字抢在了他的前面：勒维耶。

但亚当斯此刻已无暇品味自己与这位法国同行在研究课题上二度撞车的深刻寓意了，他的精力越来越多地投入到了推测未知行星轨道的计算之中。如我们在上章所说，这是一项极其复杂的计算。如果说扎克和他那些试图围捕小行星的朋友是天空警察，那么亚当斯就应该算是星探，当然不是寻找演艺明星的星探，而是星空侦探，他要做的是通过“罪犯”在“犯罪现场”，即天王星轨道留下的蛛丝马迹，来推断其行踪。

亚当斯从1780年到1840这60年的现代观测数据（其中很多是艾里提供的）中以每三年为一个间隔整理出了21组数据。他分别计算了这21组数据与波瓦德轨道的偏差，并与新行星产生的影响进行比较及拟合。由于新行星的轨道参数中除用提丢斯—波德定则确定的半长径外，其余全都是未知的，他需要通过不断调整参数来寻求最佳的拟合效果。在计算中他还采用了高斯计算谷神星轨道时所用的误差控制方法。考虑到当时

的计算主要依靠手算^{[\(1\)](#)}，这实在是一件令人望而生畏的工作。亚当斯以惊人的专注和毅力进行着计算——这也是他一贯的风格。他的兄弟乔治（George Adams）曾有一段时间陪伴他熬夜，帮他验证一些计算结果。在乔治撰写的回忆中，他提到有很多次当他实在熬不下去时，想让亚当斯去睡觉（那样他自己也可以休息），亚当斯总是说：再等一会儿。而那“一会儿”却总是无穷无尽般的漫长。在亚当斯沉醉于计算的那些日子里，他几乎神游物外，甚至在与乔治一起散步时都需要后者提醒他避开障碍物。经过这样没日没夜的努力，当下一个秋天来临时，1845年9月，亚当斯的计算终于有了结果。

注释

^{[\(1\)](#)} 当然，对数表等数学表格对部分计算可以起到辅助作用。

12 三访艾里

按照亚当斯的推算，新行星的质量约为天王星的3倍，运动轨道则是一个半长径为38.4天文单位（比一开始假定的38.8天文单位略小）的椭圆轨道。在这样一颗新行星的影响下，亚当斯将天王星的出轨幅度由原先的几十〔角〕秒压缩到了1~2〔角〕秒，并预言了新行星1845年10月1日将在天空中出现的位置。由于亚当斯在计算中只用到了现代数据，因此一个很自然、并且也很重要的问题是：他的计算是否也可以解释历史记录？为此，亚当斯进行了验证，结果发现答案是肯定的。历史记录与现代数据的跷跷板第一次被摆平了，这在很大程度上印证了计算结果的可靠性，也间接印证了新行星假设的合理性。

亚当斯这些繁复计算的完成，在时间上与他为预测新行星位置所选的1845年10月1日这一日子已相距不远。如果他能像当年的赫歇耳那样拥有一流的望远镜，且精于观测的话，完全有可能通过几周甚至——如果运气好的话——几个夜晚的观测，就能亲自发现那颗新行星，因为他所预测的位置距离新行星当时在天空中的实际位置只相差了不到两度⁽¹⁾。可惜亚当斯并没有那样的条件，于是他将自己的计算结果告诉查利斯，再次寻求后者的帮助。



查利斯也再次表现出了够意思，只不过他的“意思”似乎总也离不开书信。他接到亚当斯的请求后，于9月22日替亚当斯写了一封推荐信，让他面呈给艾里。查利斯在信中称亚当斯的计算是可以信赖的。但令人困惑的是，查利斯身为剑桥天文台的台长，自己就拥有搜索新行星所需的一切技术条件，却为何要舍近求远地把亚当斯推荐给艾里呢？而且他作为年长者，居然没有建议亚当斯正式发表那些“可以信赖的”的计算结果，这又是为什么呢？对此，一个可能的解释是查利斯其实并未真正相信亚当斯的结果。亚当斯的能力虽然已经通过替他计算彗星轨道而不止一次地得到了显现，但那些计算的难度与通过天王星轨道来反推一颗未知行星的行踪相比，无疑还相差很远。不管是出于什么原因，查利斯这位堪称当时全英国最了解亚当斯的天文学家，在这个至关重要的历史节骨眼上没有选择直接的帮助和参与，这是他与发现新行星的机会第一次擦肩而过。

9月底，亚当斯带着查利斯的推荐信来到格林威治天文台（图7）拜访艾里。这是他第一次拜访艾里，可惜艾里当时正在法国开会，让他扑了个空。出师不利的亚当斯只得留下查利斯的推荐信无功而返。艾里回到天文台后看到了查利斯的推荐信，他很快就给查利斯回了信，对错过与亚当斯的会面感到遗憾，并礼貌地表示对亚当斯的工作很感兴趣，欢迎他与自己建立通信联络。亚当斯得知这一消息后决定再次访问艾里。



图7 格林威治天文台

1845年10月21日下午3点左右，亚当斯再次来到了格林威治天文台(2)。不巧的是，艾里居然又不在。不过这次他只是暂时外出，于是亚当斯向艾里的管家表示自己过一会儿再来，并留下了一张一页纸的计算结果。亚当斯在附近溜达了大约一个小时后重新来到了艾里家。不幸的是，不知是由于管家的疏失还是其他什么原因，艾里似乎并未收到亚当斯的“拜山帖”，也不知道他会返回。因此当亚当斯第三次登门拜访时，被告知艾里正在吃午饭，不见客人(3)。因为吃午饭就不见客人，这听起来似乎有些无理，其实在英国这样一个礼仪森严的国家里却不足为奇，尤其是艾里乃是天文界的资深前辈，而亚当斯只是一位初出茅庐的年轻人，艾里在吃饭时不见亚当斯并不算出格。但尽管礼仪如此，连吃三次闭门羹还是让亚当斯失去了耐心，他没等艾里吃完午饭就返回了剑桥。

回到剑桥后，亚当斯把寻求观测支持的事搁到了一旁，他决定进一步改进自己的计算。在他第一轮的计算中，曾将未知行星的轨道半长径

假设为38.8天文单位，这是提丢斯一波德定则的预言。但提丢斯一波德定则虽已接连被天王星和小行星带的发现所支持，却终究没什么理论基础，因此亚当斯对建立在提丢斯一波德定则基础上的轨道半长径假设并不满意。在新一轮的计算中，他决定放弃这一假设，而尝试一个稍小一点的轨道半长径：37.3天文单位。

另一方面，艾里最终还是看到了亚当斯留下的那一页计算结果。两个星期之后，即11月5日，他给亚当斯回了一封信。在回信中，艾里与亚当斯一样，也质疑了提丢斯一波德定则的有效性，但他同时还提出了另外一个问题：那就是如何解释天王星出轨之谜中的径向偏差。我们在第10章中曾经介绍过，天王星轨道的径向偏差在艾里眼中是很重要的问题，他甚至认为这很可能就是解决天王星出轨问题的关键。由于他的这一看法并未得到其他天文学家的重视，因此艾里一有机会就要重提这一问题，对亚当斯自然也不例外。

但这回却轮到艾里吃闭门羹了，因为亚当斯并未对艾里姗姗来迟的信件作出回复。亚当斯的沉默落在艾里眼中无疑变成了一个信号，让他以为自己的问题已击中对方的要害，两人的联系就此中断。

那么，亚当斯为什么不回复艾里的信件呢？据他后来在一封为此事而向艾里表示歉意的信中所说，那是因为他并未意识到艾里对这一问题如此看重。很多年后，他在给一位朋友的信中则表示，他当时觉得艾里的问题太过简单，因此没有及时回复。不过亚当斯的这些解释颇有避重就轻之嫌，其可信度是值得怀疑的。艾里怎么说也是前辈，哪怕他真的提了一个太简单或不重要的问题，甚至我们把亚当斯对前不久的闭门羹一事还耿耿于怀的可能性也考虑在内，作为后辈的他似乎也没有理由用不回信这样失礼的方式来处理。这样的事情别说在英国，即便在礼仪相对宽松的其他国家，恐怕也是不合情理的。

如果亚当斯自己所说的原因不合情理，那么真正的原因会是什么呢？从逻辑上讲，最有可能的答案恐怕就是：他是因为无法及时对艾里

的问题作出明确回答，才没有回复的。这一点后来得到了一些史料的佐证，因为人们在亚当斯残存的笔记中发现他曾试图解决艾里的问题。这与他在上述信件中所说的并未意识到艾里对这一问题的看重，以及认为艾里的问题太过简单显然是有些自相矛盾的。

但无论出于什么原因，忽视也好、为难也罢，甚或只是纯粹的偶然，亚当斯与艾里三番两次无法建立有效的联系，无论对他们两人，还是对整个英国天文学界都是一个极大的遗憾。就在机遇从亚当斯、查利斯和艾里的指缝间一次次遗落的时候，一位法国天文学家把自己的目光投向了天王星的出轨问题。

这已是此人第三次与亚当斯在相同的问题上相遇。

注释

[\[1\]](#) 需要提醒读者的是，这一偏差是指计算位置与实际位置的偏差，而非计算误差。（请读者想一想，这两者的差别是什么？）后来有人对这一数据，乃至英国方面的整个故事都提出了质疑，对此我们将在后文中加以介绍。

[\[2\]](#) 据说人们并未在当时遗留的日记、信件等文字记录中找到亚当斯第二次访问艾里的确切日期，因此10月21日这个日期是后人的推断。

[\[3\]](#) 关于这一点，艾里夫人曾留下过两个不同版本的书面说法，后人据此认为有关艾里一家当时正在吃午饭的传言未必确凿。由于艾里一家当时正在做什么对整个事件的发展并无特殊重要性，因此本文不予细究，这里提一下只是为了告诉读者史学界对这一细节存有不同看法。

13 殊途同归

这位法国天文学家的名字大家一定猜出来了。是的，他就是两次在彗星轨道计算中与亚当斯不期而遇的勒维耶。勒维耶有着与亚当斯同样杰出的数学技能，不过他的天文之路却略显曲折。1830年，初出茅庐的勒维耶在报考法国第一流的理工学校巴黎综合理工学院（Ecole Polytechnique）的竞争中不幸落败。由于勒维耶在当地学校的成绩一向十分优异，他父亲将失败的原因归咎于当地整体教育水平的低下。望子成龙的他毅然变卖了房产，将勒维耶送到巴黎复习备考。第二年，脱离了山沟沟的勒维耶终于变成了金凤凰，不负众望地进入了巴黎综合理工学院。

与亚当斯一样，勒维耶以最优异的成绩通过了学校的毕业考试。不过毕业后的勒维耶却一度进入了与天文学风马牛不相及的政府烟草部门，并跟随化学家盖-吕萨克（Joseph Louis Gay-Lussac）从事过一些化学方面的研究^{[\(1\)](#)}。1837年，当巴黎综合理工学院的一个天文学教职出现空缺时，盖-吕萨克建议并推荐勒维耶获得了这一教职。虽然对导师建议的转行感到意外，但勒维耶很快就发现天文学是一个可以充分展现自己数学才华的迷人领域。转行天文后的勒维耶主要从事天体轨道的计算与分析。短短几年间，他便在该领域树立起了自己的名声。

勒维耶的理论研究有着鲜明的系统性，这一点与当年赫歇耳的观测工作颇有异曲同工之处。自1840年以来，勒维耶对太阳系天体的运动做了近乎地毯式的研究，先后考察了水星、金星、地球、火星、木星、土星及若干彗星的运动。1845年秋天，在巴黎天文台台长阿拉果（FranÇois Arago）的提议下，他将注意力转向了当时已知的最后一个行星：天王星。

与亚当斯一样，初涉天王星问题的勒维耶也对波瓦德轨道进行了细

致分析，也发现并纠正了一些错误，他的结论也和亚当斯一样，那就是波瓦德轨道已经无可救药了——仅凭木星和土星的影响是无论如何也摆不平天王星轨道的。接下来，勒维耶又近乎地毯式地逐一分析了我们在第9章中介绍过的几种试图解决天王星出轨之谜的假设，并将它们一一排除。这样，他顺理成章地将注意力转向了当时已知的最后一个假设：新行星假设，并且与亚当斯一样，走上了用纸和笔寻找新行星的艰难之旅。

作为计算的出发点，勒维耶也采用了提丢斯—波德定则，把新行星的轨道半径假定为38.8天文单位（在计算过程中微调为38.4天文单位，与亚当斯第一次计算的结果相同）。不过与亚当斯所用的椭圆轨道不同的是，他假定新行星的轨道为圆形。为了确定新行星在轨道上的位置，他将轨道按角度均匀地分割成了40个区段，每段 9° （因为整个圆周有 360° ）。显然，在任何一个选定的时刻——勒维耶将之选为1800年1月1日——新行星必定位于这40个区段中的某一个区段内。那么它究竟位于哪一个区段呢？勒维耶再次发挥了自己的系统风格，他逐一考察了新行星在选定时刻位于40个区段中的任何一个区段内所能对天王星轨道产生的影响。通过极其繁复的计算、对比和排除，到了1846年5月底，勒维耶终于找到了能够使天王星轨道最接近观测结果的那个区段。在此基础上，他预言了1847年1月1日新行星所处的位置。

勒维耶的计算结果与亚当斯的相当接近。英吉利海峡两边的这两位数学高手的智慧之剑指向了同一个天区，只不过当时勒维耶和亚当斯谁也不知道对方的工作。

与亚当斯不同的是，勒维耶公开发表了自己的计算，从而引起了一定程度的关注，因为那时天王星出轨之谜已经困扰天文学家们达半个世纪之久，新行星假设成为解决这一谜团的主流假设也已有差不多十个年头，这还是第一次有人计算出新行星的确切位置（亚当斯的结果因为没有发表，除查利斯和艾里外，尚处于无人知晓的状态）。1846年6月下

旬，勒维耶的论文抵达了艾里所在的格林威治天文台。

如果说其他天文学家对勒维耶结果的关注在很大程度上是出于新奇，那么对艾里来说，勒维耶的结果则带来了一定的震动，因为这一结果与他大半年前曾经见过的亚当斯的结果相当接近⁽²⁾。亚当斯在当时还是一个籍籍无名的年轻人，而勒维耶已有一定的知名度，艾里也许可以忽略亚当斯，但对勒维耶的结果却不能等闲视之。更重要的是，在这么困难的问题上，两个人同时算错并非不可能，但错得如此接近却令人难以置信。因此，这时的艾里对亚当斯和勒维耶的结果已不得不刮目相看，他甚至向包括小赫歇耳（John Herschel，发现天王星的老赫歇耳的儿子）在内的几位朋友及同事提及了这两人的计算彼此接近，以及在近期内借助计算结果发现新行星的可能性。

不过，要让艾里信服勒维耶的计算还必须解决他心头的一个老大难问题，那就是他当年曾问过亚当斯，却未得到回答的那个天王星轨道的径向偏差问题。这一问题依然盘亘在艾里的脑海里，于是他写信给勒维耶，询问他的计算能否解决这一问题。1846年7月1日，艾里从勒维耶的回信中得到了非常肯定的答复。这下艾里终于信服了。即便如此，他却并未采取立即行动。从我们这个史话的角度看，艾里此时的迟钝是一件非常奇怪的事情，但我们不能忘记，在这位皇家天文学家的日程中本就有太多的事情需要他去关注，虽然那些事情的重要性在事后看来与他错过的事情相比根本就不值一提。幸运的是，艾里早年的一位数学老师在关键时候击碎了他的迟钝。7月6日，艾里与这位名叫皮考克（George Peacock）的数学教授谈及了天王星出轨问题及亚当斯和勒维耶的计算。老教授对艾里的迟钝大为惊讶，敦促他立即采取行动。



三天后，艾里终于采取了行动。

而这时候，勒维耶在做什么呢？他和亚当斯一样，投入到了新一轮的精密计算之中。在这一轮的计算中，他决定放弃前一轮计算所采用的两个不太令人满意的假设：其中一个提丢斯-波德定则，勒维耶和亚当斯一样，认为这是一个没有足够理论基础的假设；另一个则是圆轨道假设，这无疑是一个过于特殊的假设。放弃这两个假设后，勒维耶将新行星的轨道调整为了半长径为36.2天文单位的椭圆轨道。

勒维耶对新计算的沉醉，在无意间为艾里及英国天文学界创造了一个难得的机会。因为一方面，沉醉于计算的勒维耶把寻求观测支持的事情搁到了一边；另一方面，勒维耶进行新计算这一消息本身在一定程度上降低了欧洲大陆的天文学家们对他前一轮计算的重视程度。这样的局面对于艾里以及他的少数英国同事来说无疑是非常有利的，因为只有他们知道亚当斯的结果，从而也只有他们才知道勒维耶的计算并非孤立结果。一个复杂的计算，它是孤立结果还是得到过独立来源的佐证，其可信度是截然不同的。英国人曾将亚当斯提供的先机轻轻葬送，但此刻的他们趁着欧洲大陆的天文学家们对勒维耶的计算将信将疑，心存观望之际，提前洞悉了这一结果的可信度，并决定展开行动，这无疑是再次将先机揽到了自己身旁^[3]。

那么，英国绅士们能够把握住这稍纵即逝的先机吗？

注释

^[1] 盖-吕萨克在化学方面有不少贡献，比如我们在中学化学课上接触过的气体化合体积定律，即盖-吕萨克定律，就是以他的名字命名的。

^[2] 如果将他们的计算统一折算成平均日面经度（mean helio longitude）的话，那么亚当斯的结果是1845年10月1日新行星位于经度323.5°；勒维耶的结果则是1847年1月1日新行星位于经度325°。

[\(3\)](#)严格地讲，欧洲天文学界并非铁板一块，在欧洲的某些地方曾有过一些零星的观测。

14 剑桥梦碎

如果要在格林威治天文台的历任台长中评选几位从事天文观测最少的台长，艾里无疑会名列前茅。自从1835年出任台长以来，8年的时间里，艾里参与过的观测仅占同期天文台观测总数的0.2%。即便在发现新行星的荣誉有可能唾手而得的时候，艾里仍没有打算亲自参与观测。他更感兴趣的问题倒是让谁来摘取这一荣誉。结果他选择了剑桥天文台，那是他就任格林威治天文台台长之前任职过的地方，那里有他亲自督建的高品质的诺森伯兰望远镜（Northumberland telescope）[\(1\)](#)（图8）。而且，亚当斯、艾里自己以及剑桥天文台的现任台长查利斯都是剑桥的毕业生，让剑桥天文台成为新行星的发现地，无疑可以演绎一出最完美的“剑桥天文故事”。



图8 安放诺森伯兰望远镜的圆屋顶

不过平心而论，要说艾里选择剑桥天文台而非自己所在的格林威治天文台是纯粹的心血来潮或浪漫胸怀，却也并非实情。事实上，格林威治天文台的名头虽大，可是由于承担了太多时间及经纬方面的测定任务，它所拥有的望远镜已经按这些特殊任务的需要进行了改动，比方说它的方向已被固定在了特定的子午线（即经线）上，以便能精确测定日月星辰穿越子午线的时间，而且它的放大倍率也比不上剑桥天文台的望远镜（格林威治天文台的望远镜口径只有6.7英寸，而剑桥的诺森伯兰望远镜的口径达11.75英寸）。这些都使得格林威治天文台已变得不再适合行星搜索工作了。因此艾里的选择也可以说是形势使然。

主意既定，艾里便于1846年7月9日写信给查利斯，叙述了剑桥天文

台搜索新行星的有利条件，然后请他展开搜索。艾里表示，如果查利斯本人没有时间的话， he 可以从格林威治天文台抽调一位助理予以协助。但艾里的信发出之后却变成了泥牛入海，查利斯并未及时回复。等了四天没有回信，艾里终于着急了，他再次写信给查利斯，提醒他寻找新行星的重要性应当凌驾于任何不会因推延而失效的其他工作之上。

查利斯居然还是没回信。

原来艾里的这位继任者当时并不在剑桥，而是在度假。当年的天文学家既没有电话也没有电子邮件，更没有个人博客可以随时向外界展示自己的行踪。艾里对查利斯度假的消息一无所知，白白着了一场急。7月18日，查利斯终于回到了剑桥天文台，他立刻给艾里回了信，表示将尽快展开观测。艾里随即给查利斯提供了一个以勒维耶和亚当斯的计算结果为中心，东西范围 30° ，南北范围 10° 的区域作为搜索范围。

但查利斯在动作迟缓方面并不比他的前任艾里先前的拖拉来得逊色，他的“尽快”足足经过了十天的时间才付诸实施。在那期间，他向亚当斯提及了自己将要搜索新行星的消息。此时亚当斯的新一轮计算已接近完成，他向查利斯提供了一些新的数据⁽²⁾。这时距离艾里收到勒维耶的回信已相隔近一个月，所幸欧洲大陆的情势并无实质变化，勒维耶的新一轮计算仍未结束，欧洲大陆的各主要天文台也仍无动于衷。

7月29日晚，查利斯的搜索行动正式展开。英国天文界的成败在此一举。

按照后来查利斯向艾里及英国报刊提供的叙述，在搜索中，他首先以亚当斯提供的位置为中心，观测了宽度为9〔角〕分（1〔角〕分等于 $1/60$ 度）的区域中所有视星等在11以上的天体。几天之后他的观测因天气而受阻。8月12日天气转好，查利斯对7月30日曾经观测过的天区进行了复测。然后他开始对比7月30日的观测结果与8月12日的复测结果。这种对比是搜索行星的标准手段，如果在对比中发现任何一个天体的位置发生了变化，那么这个天体就有可能是查利斯所要寻找的新行星。一

组、两组、三组.....查利斯一连对比了39组数据，全都匹配得完美无缺，这表明那些都不是他要寻找的新行星。虽然还剩下一些数据尚未对比，但查利斯觉得这一天的对比不会有什么收获了。他想起自己手头还有一些彗星数据需要处理，于是便提前终止了对比工作。

这一决定酿成了查利斯一生最大的遗憾，也彻底葬送了艾里梦想的“剑桥天文故事”。



查利斯完全没有想到，在这场无形的竞争中，就在他迎来长久阴霾之后的第一个好天气时，幸运女神又一次——也是最后一次——将垂青的目光投到了英国人的头上。新行星的数据此刻就静静地躺在他8月12日的复测记录中。那一天，查利斯只要再多对比10组数据，就会发现8月12日所记录的第49个天体——一颗蓝色的8等星——在7月30日的记录中是不存在的。这说明那个天体7月30日还不在他所观测的天区中，8月12日却进入了该区域，那是一个移动的天体，那个移动的天体正是艾里要他寻找的新行星！

一招失误，满盘皆输。

在8月份余下的日子里，查利斯继续对附近的天区进行搜索，结果一无所获。9月初，他放弃了搜索。

注释

[\[1\]](#) 该望远镜是一位诺森伯兰公爵（Duke of Northumberland）于1833年捐助的，故而得名。

[\[2\]](#) 关于亚当斯向查利斯提供的究竟是什么数据，后来有人提出了质疑。质疑者认为亚当斯当时提供的其实是勒维耶第一轮计算的结果。如果那样的话，那些数据与亚当斯当时即将完成的计算应该没什么关系。

15 欲迎还拒

查利斯的失败，宣告了英国人在这场几度领先的无形竞争中丧尽先机，黯然出局。虽然此刻他们还不清楚自己究竟失去了什么，但历史的风标已无可阻挡地偏向了后来居上的欧洲大陆。

就在查利斯终止新行星搜索之前不久，1846年8月底，勒维耶完成了他的新一轮计算。那时候，他在整个计算中用去的稿纸数量已经超过了10000页。勒维耶的新结果与原先的结果相当接近（偏差只有 1.5° ），这是一个好兆头，它表明勒维耶的计算方法很可能具有良好的稳定性。按照勒维耶的计算，虽然新行星的轨道半径比天王星大了将近一倍，但由于其质量也比天王星大得多，因此亮度依然可观。勒维耶很清楚，再好的计算若是离开了观测的验证，也只能是空中楼阁。他已经用了一年的时间来构建这座宏伟的楼阁，现在是该让一切落地生根的时候了。于是他开始尽其所能地劝说天文学家们对新行星展开搜索。

而此时的英国几乎只剩一个人还在为新行星的命运做最后的奔走，他就是亚当斯。比勒维耶晚了几天，亚当斯也完成了自己的新计算。只不过，勒维耶循正常而公开的学术渠道发表了自己的所有计算，而亚当斯却仍在继续那种曾让自己一再碰壁的私下交流。9月2日，他给艾里写了一封信，一来通报自己的第二轮计算结果，二来则答复一年前艾里在信中问起过的天王星轨道径向偏差问题（这封信从一个侧面说明亚当斯当年对径向偏差问题的沉默，并非是因为没有意识到艾里对这一问题的看重，或觉得该问题太简单）。可惜的是，亚当斯和艾里的每一次重要交往似乎都注定要以失败告终。亚当斯的信件抵达格林威治天文台时，艾里又出了远门。不过这回亚当斯多少也学了一点乖，不再把宝完全押在艾里一个人身上了，他决定赶往英国科学进步协会（British Association for the Advancement of Science）在南安普敦（Southampton）

的一次会议，以便报告自己的结果。

可人要是背了运，喝凉水都会塞牙。

当亚当斯赶到南安普敦时，天文方面的会议早已结束。错过了会期的亚当斯只能郁闷地再次将自己埋首于计算之中，他将新行星的轨道半径进一步缩小为34.4天文单位，开始了自己的第三轮轨道计算。

另一方面，勒维耶的命运虽然比亚当斯顺利，除了艾里等少数人外，不仅整个欧洲都将他视为新行星位置的唯一预言者，他的工作甚至还远隔重洋传到了美国。自8月份以来，勒维耶预言新行星位置的消息更是越出了学术界的范围，得到了媒体的宣传。可当他试图说服各大天文台将那些溢美之词，以及对新行星的期盼之意转变为货真价实的搜索行动时，却遭遇了意想不到的困难。各天文学台的“老总”们虽毫不掩饰对他结果的极大兴趣，以及对他水平的高度赞许，可一涉及动用自己手下的人力和设备进行新行星搜索时，却一个个支支吾吾、推三阻四起来。甚至连他的本国同行也不例外，一年前亲自敦促他研究天王星出轨问题的巴黎天文台台长阿拉果只进行了极短时间的搜索就放弃了。

读者也许觉得奇怪，发现新行星是何等的美事？各大天文台应该争先恐后，抢破脑袋才是，怎么反倒你推我让，欲迎还拒呢？难不成是“老总”们突击学习了“孔融让梨”的先进事迹？各大天文台之所以会有这样奇怪的反应，主要有两大原因：第一是信心不足。谁都知道发现新行星意味着巨大的荣誉，但同时也都知道寻找新行星是一件很困难的事情。虽说此次的情况有所不同，勒维耶已经计算出了新行星的位置，而且轰传天下。可这计算新行星位置的壮举，乃是前所未闻的故事，天文学家们心里究竟信了几成，恐怕只有他们自己才清楚。溢美之词是廉价的，观测时间却是无价的，该不该用无价的时间去验证廉价的评语，这是让“老总”们不无踌躇的事情。

第二个原因则是制度死板。当时各大天文台都有繁重的观测任务，也都有比较死板的规章制度，对观测日程做哪怕细微的变更都不太容

易，要想凭空插入一个耗时未知，结果莫测的行星搜索计划更是难上加难。格林威治天文台甚至还有过观测助理因擅自寻找新行星而受到艾里惩罚的事情。因此即便像查利斯那样既得到艾里的嘱托，又有台长的权力，并且因知晓亚当斯与勒维耶的双重结果而具备信心优势的人，也但愿花费很有限的时间和精力进行观测，且还心猿意马、草率从事，以至于功败垂成。而艾里本人把观测任务交给剑桥天文台，虽有演绎剑桥故事的美好心愿及其他客观原因，但心底里——据后人分析——也是不想打乱格林威治天文台的正常工作。

16 生日之夜

一次次客气的回绝让勒维耶很是沮丧，他搜肠刮肚地寻找关系，试图找到一个突破口。这时他忽然想起了自己曾经认识过的一位柏林天文台的天文学家，此人名叫伽勒（Johann Gottfried Galle），是柏林天文台台长恩克（Johann Franz Encke）的助理。说起来，勒维耶与伽勒的关系其实疏远得很，唯一值得一提的联系是一年前伽勒曾给勒维耶寄过一份自己的博士论文。而忙于计算的勒维耶甚至连封感谢信都没有回。真所谓此一时彼一时也，若非如今这档子事，勒维耶这辈子能否想得起伽勒来都是个问题，而此刻勒维耶一想到伽勒就觉得亲切无比，犹如看到了救命稻草。1846年9月18日，勒维耶给伽勒写了一封信，将伽勒一年前的博士论文狠狠地夸奖了一通，然后笔锋一转，谈到了自己预言的新行星位置，他希望伽勒能帮助寻找这颗行星。



德国天文学家伽勒 （1812-1910）

9月23日，勒维耶的信送到了伽勒手中。



虽然被冷落了一年，能够收到当时已颇有名气的勒维耶的来信（而且还是满含赞许的来信），伽勒还是感到非常兴奋，并且他也被勒维耶的预言深深吸引了。勒维耶的信终于落到了能被它打动的人手里，不过更妙的则是这封信的抵达时间：9月23日，这一天正好是伽勒的老板恩克台长的55岁生日。伽勒虽是柏林天文台的资深成员，但按规矩却没有擅自使用天文台的望远镜进行计划外观测的权力，他想要观测新行星，必须得到台长恩克的允许。

恩克作为台长，消息自然是灵通的，他早就知道勒维耶预言新行星的事，但和其他天文台的台长一样，他对此事也一直采取了旁观的态度。换作平时，伽勒的要求可不是那么容易过关的。不过人在生日的时候心情通常比较愉快，而且那天晚上同事们早已约定在恩克家中庆祝他的生日，并无使用望远镜的计划，因此在伽勒的软磨硬泡之下，恩克终于答应给对方一个晚上的时间进行观测。



德国天文学家达雷斯特 （1822-1875）

拿到了尚方宝剑，伽勒拔腿就往观测台走。这时一位年轻人叫住了

他。此人名叫达雷斯特（Heinrich Louis d'Arrest），当时还是柏林天文台的一位学生，他碰巧旁听到了伽勒与恩克的谈话。达雷斯特请求伽勒允许自己也参加观测。由于天文观测不仅是观测，而且还需要进行数据的记录与比对，有助手参与显然是非常有利的，于是伽勒答应了达雷斯特的请求，两人一同前往观测台。

我们在第4章中曾经介绍过，发现行星的主要途径有两种：一种是通过行星的运动（比如小行星的发现），另一种则是通过行星的圆面（比如赫歇耳发现天王星）。由于通过运动发现行星通常需要对不同夜晚的观测数据进行对比，而恩克只给了他们一个夜晚的时间，因此伽勒和达雷斯特将希望寄托在了观测新行星的圆面上。他们将望远镜指向了勒维耶预言的位置，以那里为中心展开了观测。

那个夜晚秋高气爽，万里无云，是进行天文观测的绝佳天气。但天气虽然帮忙，运气却似乎并不垂青于他们。时间一分一秒地过去，他们并未发现任何显示出圆面的天体。夜色越来越浓，希望却越来越淡，难道勒维耶的预言错了？又或是预言没错，但误差太大，从而新行星离预言的位置太远？如果是这样，他们就必须扩大搜索范围，而这显然不是短短一个夜晚就能搞定的。

百般无奈之下，达雷斯特提议了一个方法：他们虽然只有一个夜晚的观测时间，从而不可能通过对自己的数据进行对比来发现新行星的运动，但他们搜索的这片天区以前也有人观测过（虽然目的各不相同）。如果他们刚才观测过的天体中有一颗是行星，那么在人们以前绘制的星图上，它显然不会处在同样的位置，甚至应该完全不在同一片天区里，因为以前绘制的星图与他们自己的观测在时间上相距较远。

由此看来，只要他们能在自己的观测中发现一颗不在星图上的天体，那个天体极有可能就是他们想要寻找的新行星。这是一个绝妙的新思路。当然，他们的运气好坏还取决于星图的详尽程度。

仿佛与他们的机智遥相呼应，柏林天文台（图9）最近恰好编过一

份详尽的星图，那份星图此刻就放在恩克的抽屉里。伽勒和达雷斯特赶紧找来了那份星图，然后由伽勒将望远镜中看到的天体的位置一个个报出来，达雷斯特则在星图上一一寻找——找到一个就排除一个。半个小时过去了，兴奋的时刻终于来临，当伽勒报到一颗视星等为8等，与勒维耶预言的位置相差不到1度的暗淡天体时，达雷斯特喊了起来：那颗星星不在星图上！

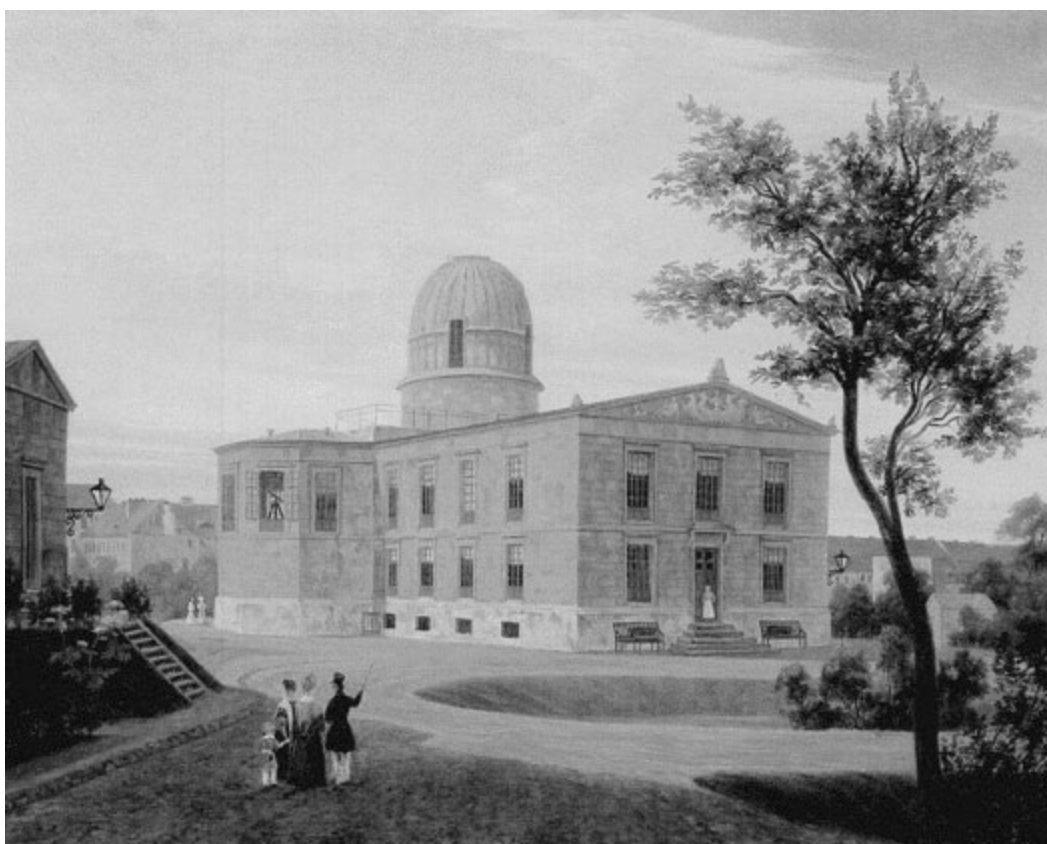


图9 柏林天文台

此刻的时钟已悄然划过零点，崭新的一天已经来临⁽¹⁾。在这个不眠之夜里，一个天体力学的神话已被缔造，天文学的历史翻开了辉煌的一页。此时恩克的生日派对仍在进行，激动不已的伽勒和达雷斯特赶到恩克的住所，向他报告了这一消息。恩克立即中断了生日派对，与他们一同赶往观测台，三人一直观测到黎明。第二天，在同样完美的天气条

件下，他们又仔细复核了一次，发现那个天体的位置移动了，并且移动的幅度与勒维耶的计算完全吻合。毫无疑问，他们已经发现了勒维耶预言的新行星^{[\(2\)](#)}。

9月25日早晨，走下观测台的伽勒写信向勒维耶报告了发现新行星的消息。

这个消息很快就席卷了整个天文学界，并将在不久之后掀起一场风暴。

注释

^{[\(1\)](#)} 尽管如此，人们通常仍将1846年9月23日作为新行星的发现日。

^{[\(2\)](#)} 经过仔细的观测，他们也确定了新行星的圆面大小，比勒维耶预言的小了20%左右。

17 名动天下

虽然近代天体力学史上从来就不乏精密的计算和预言，比如我们在第6章中曾经提到，高斯预言的谷神星位置与实际观测只差0.5度。至于有关日食、月食及彗星周期等的预言，则更比比皆是。但那些计算所涉及的天体，其存在性及部分轨道数据都是已知的，所有的计算和预言都只是从有关该天体的已知数据出发，来推测未知数据。而像勒维耶这样通过已知行星的运动，来间接推算一颗远在几十亿千米之外，没有任何观测数据的未知行星的轨道，并将其位置确定到如此精密的程度，这不仅是前所未有的壮举，而且充满了引人遐想的空间。一时间，所有人都被这令人炫目的伟大成就所震撼，这一成就的“总设计师”勒维耶几乎在一夜之间就达到了自己一生荣耀的顶点。来自欧洲各地的赞美与祝贺如雪片般飞来，很多人激动地将勒维耶的成就称为天文史上最伟大的成就。

在伽勒给勒维耶报信的同时，他的老板恩克也亲自给勒维耶写了信，在信中，除了向勒维耶表示“最诚挚的祝贺”外，恩克还写道：“您的名字将永远与对万有引力定律有效性的能够想象得到的最惊人验证联系在一起”。德国天文学家舒马赫（Heinrich Schumacher）的评论则是：“这是我所知道的最高贵的理论成就”，这位舒马赫曾试图帮助勒维耶联络德国及英国的天文学家进行新行星搜索，可惜那些被他联络到的天文学家无一例外地丧失了机会。除天文学界外，欧洲的媒体也迅速报道了这一消息，并在公众中激起了极大的兴趣。10月5日，新行星发现后的第十天，法国科学院每周一次的例行会议几乎成了勒维耶的明星秀，闻讯而来的民众把科学院的入口挤得水泄不通，人人争睹勒维耶的巨星风采，每张嘴都在念叨着勒维耶的名字。甚至连法国国王也被勒维耶的盛名惊动，亲自聆听了勒维耶对自己发现的介绍。



在这涌动的热潮中，许多法国民众开始将新行星称为“勒维耶星”。提议以发现者的名字命名行星，这在行星发现史上并非头一遭，在法国尤其如此。当年赫歇耳发现的天王星在法国就一度被称为“赫歇耳星”，更何况此次发现新行星的首要功臣就是法国人。这时候，倒是勒维耶本人很谦虚地提议了一个不同的名字：奈普顿（Neptune），这是罗马神话中的海洋之神。这个名字既符合行星命名的神话惯例，又与新行星在望远镜里呈现的美丽蓝色珠联璧合，是一个很漂亮的提议⁽¹⁾。不过这一名字尚未得到公认，连续几天的“群众运动”及“勒维耶星”的“黄袍加身”就使勒维耶的想法产生了变化。他觉得新行星若果真被命名为“勒维耶星”，倒也是一件很幸福的事情。这样的命名虽有违惯例，但考虑到此次的情形是如此的独一无二，勒维耶觉得自己享受一个独一无二的命名也并不为过。在他的示意下，巴黎天文台的台长阿拉果公开提议将新行星命名为“勒维耶星”。但这一提议终究没能与已成主流的神话命名体系相抗衡，更何况此时此刻，一场巨大的风暴已然来临，小小的命名之争很快就淹没在了惊涛骇浪之中。等到风浪平息之后，最终还是海洋之神成为了新行星的名字，在中文中，这一行星被称为海王星⁽²⁾。

海王星被发现时的视星等为8，虽然超出了肉眼所能辨别的极限，但在望远镜所能观测的天体中却是比较亮的。因此与天王星的情形一样，天文学家们很快就发现海王星其实也早在其被伽勒和达雷斯发现之前，就已被反复记录过。这其中最该痛哭流涕的无疑是查利斯，在与新行星擦肩而过的所有人中，他是唯一一位以搜寻新行星为目的，并且观测到了目标，却仍失之交臂的人。悔恨排行榜上的亚军则属于法国天文学家莱兰德（Michel Lalande），此人的“冤情”堪比其同胞拉莫尼亚（参阅第4章）。1795年5月8日及5月10日，莱兰德两次观测到了海王星。次数虽不算多，但与拉莫尼亚不同的是，莱兰德明确注意到了该天体在两天之中的位置变化。此时此刻，新行星的发现已呼之欲出。但

令人难以置信的是，莱兰德竟鬼使神差般地认定自己5月8日的观测是不准确的，而且连进一步的确认及后续观测都没做，就将这千载难逢的机会拱手送还给了命运女神，从而创下了行星观测史上最离奇的失误。

在曾经记录过新行星的人之中，最让人意想不到的则是伽利略。1979年，人们发现这位科学启蒙时代的宗匠竟然早在1612-1613年间——即不仅比海王星的发现早了两百三十多年，甚至比天王星的发现还早一百七十多年——就至少两次观测到了海王星。另外值得一提的是，小赫歇耳曾在1830年的一次天文观测中搜索过距离海王星当时的位置只差0.5度的天区。小赫歇耳很好地继承了父亲的事业，当时已成为英国最有声望的天文学家之一。以他的观测设备及观测水平，若当时他的观测区域稍稍扩大一点，就极有可能缔造一段父子双双发现新行星的佳话。但这样的佳话假如出现，勒维耶用笔尖发现海王星的更伟大的奇迹将不复存在。小赫歇耳在给朋友的信中表示，如果那样的话，连他自己都将感到遗憾。这句话也许是心里话，也许只是一种风度，但对于行星发现史来说，这句话倒是千真万确的。海王星以如今这种方式被发现，实在是行星发现史上最动人的故事。

不过这故事虽然动人，却也没有后人渲染的那样完美，这是后话。

注释

[\[1\]](#) 有关这一提议的由来，勒维耶在给伽勒的一封信中声称是法国经度局（The Bureau of Longitude）的决定，但法国经度局却否认了这一说法。人们一般认为，这一命名是勒维耶自己的提议，至多曾与经度局的人有过非正式的交流。

[\[2\]](#) 海王星这一名称直到1847年才基本得到公认，但为了方便起见，我们在下文讲述1847年以前的事件时也将用这一名称来称呼新行星。

18 轩然大波

海王星的发现在知道亚当斯工作的一小部分英国天文学家中引起了极大的震动。发现海王星的消息传到英国时，艾里正在欧洲大陆度假，当时在英国的知情人除亚当斯本人外，主要有两个：一个是查利斯，另一个则是小赫歇耳。

小赫歇耳成为知情人的具体时间史学界尚有争议，传统的说法是他曾在6月29日皇家天文台的一次会议期间听艾里提到过亚当斯与勒维耶的计算^{[\(1\)](#)}。那是艾里极罕见的一次向他人提及亚当斯的名字，那个消息给小赫歇耳留下了深刻印象。9月10日，他在英国科学进步协会的一次演讲中，将预言海王星的位置比喻为哥伦布从西班牙海岸直接看到美洲^{[\(2\)](#)}。当时海王星尚未被发现，小赫歇耳并未在这番泛泛之语中提及预言者的名字，不过由于勒维耶的工作早已广为人知，几乎所有的听众都以为小赫歇耳指的就是勒维耶的预言。现在海王星已被发现，勒维耶也已名动天下，作为英国天文界的领军人物之一，小赫歇耳不愿坐视英国在这场无形竞争中一败涂地。10月3日，他在伦敦的一份周报上发表文章，公布了亚当斯在整个事件中的角色。这是这一事件的英国版首次被公开。小赫歇耳在文章中除了提及亚当斯的结果外，还表示正是因为知道亚当斯与勒维耶的结果几乎相同，才使他有足够的信心将预言海王星的位置比喻为哥伦布从西班牙海岸直接看到美洲^{[\(3\)](#)}。

在除亚当斯本人之外的三位英国知情人中，小赫歇耳无疑是最没有心理包袱的，因为他在这一事件中纯粹是旁观者。与他不同的是，艾里与查利斯很早就知道了亚当斯的结果，因此这两人对英国在这一竞争中的落败很难脱得了干系。尤其是查利斯，他的疏失对于一位职业天文学家来说堪称是丑闻。查利斯是9月30日得知海王星被发现的消息的，当时他还不知道自己早在一个多月前的8月12日就曾观测到过海王星，因

此心中尚无愧意。不仅没有愧意，他还有苦水要倒。因为他在9月29日看到了勒维耶发表的最新计算结果，那篇文章重新引起了他对新行星的兴趣，当天晚上，他恢复了已中断近一个月的搜索，并且成功地发现了一个有圆面的天体——那正是海王星。可惜没等他有机会复核，就传来了海王星已被发现的消息。查利斯觉得自己实在有点冤，运气也实在有点背，因此他立即给剑桥的一份刊物写了信，除提及亚当斯的工作外，还着重提到自己过去两个月以来一直在从事着早晚会成功的搜索工作，并在9月29日事实上独立地发现了海王星。查利斯的信也发表于10月3日。

若干天之后，当查利斯发现自己一个多月前的重大疏失时，他的自我惋惜才转变为悔恨与惭愧。

10月11日，艾里回到了英国，他在9月29日就得知了海王星被发现的消息。无论从学术地位还是实际作用而论，艾里在整个英国版的故事中都处于中枢地位，他很快也对事件作出了反应。不过，他没有诉诸媒体，而是直接给勒维耶写了信。在信中艾里告诉勒维耶，英国方面在他之前就有过完全相同的预言。虽然艾里表示自己这封信的目的绝不是要抹杀勒维耶的贡献，并且他也承认英国方面的工作不如勒维耶的工作来得广泛，但他对“时间上更早”及“结果相同”这两点的强调，还是让勒维耶很受伤。

在英国方面的主要当事人中，唯一未发表声明的是亚当斯本人，他虽然很沮丧，但没有参与优先权之争。相反，他将精力投入到了利用已经公布的观测数据计算海王星的真实轨道上来，并于10月份完成了计算，成为最早在直接观测数据之上完成海王星轨道计算的天文学家。

勒维耶收到艾里的来信时，小赫歇耳和查利斯的文章也几乎同时传到了法国。这突如其来的三柄利刃让勒维耶既感震惊又觉震怒。勒维耶的震惊和震怒是有道理的，艾里在海王星发现之前与他有过多次信件往来，如果英国方面早就有过与他平行的工作，艾里为什么早不提晚不

提，偏偏要等到海王星被发现之后才提？查利斯的举止更是可疑，他在刊物上的声明发表之后，又于10月5日给不止一位欧洲大陆的天文学家去信，讲述自己9月29日发现却来不及确认海王星的“祥林嫂”故事。而在那些故事中他却只字未提亚当斯的名字，这不是前后矛盾又是什么？至于小赫歇耳，他竟然声称对勒维耶结果的信心乃是因为其与名不见经传的亚当斯的结果相吻合，这对勒维耶来说简直太伤自尊了。

来自英国方面的消息不仅激怒了勒维耶，也激怒了整个法国天文界。在他们看来，这分明是英国方面蓄意捏造事实，企图抢夺荣誉的卑劣行径。人不可以无耻到这种地步，法国天文学家们的心里，那是相当的愤怒，他们立即展开了犀利的反击。10月19日，巴黎天文台台长阿拉果在巴黎科学院的会议上发表了声援勒维耶，讨伐艾里、小赫歇耳及查利斯的檄文。在这篇檄文中，阿拉果大量援引了艾里等人写给法国天文学家的信件，指出其相互矛盾之处，并痛斥英国人的卑劣企图。阿拉果在檄文的最后情绪激昂地指出：在每一双公正的眼睛里，这一发现都仍将是法国科学院的辉煌成就，也将是让子孙后代景仰的最高贵的法国荣誉。

阿拉果的檄文发表后，法国乃至欧洲其他国家的媒体都迅速跟进，展开了对艾里等人的围剿。法国的有些报道干脆将这三人称为“窃星大盗”（考虑到海王星的大小，这罪名在理论上可比地球上的“窃国大盗”大得多）。更糟糕的是，阿拉果所引的艾里等人与法国同行的通信一经曝光，在英国天文学界也引起了轩然大波。因为艾里与查利斯不仅从未向法国同行们提及过亚当斯的工作，也向绝大多数英国同行隐瞒了消息。这一点让许多英国天文学家也感到了愤怒，这其中有位天文学家叫作辛德（John Russell Hind），他曾在格林威治天文台当过助理。伽勒发现海王星的消息传到英国后，他是第一位重复这一发现的英国人。但在那之前，他就曾与查利斯讨论过搜索新行星的问题。倘若查利斯未曾向他隐瞒亚当斯的工作，他也许早就展开了认真的搜索。而如果艾里

与查利斯及早向英国天文界全面报告亚当斯的工作，说不定会有更多的英国天文学家投入搜索行动。不仅如此，更有人指出，倘若艾里与查利斯在1845年秋天亚当斯的第一轮结果出来之后就认真对待，则历史说不定早已被改写，根本就没法国人什么事。从这个意义上讲，艾里与查利斯是导致英国天文学界整体失利的罪魁祸首。某些激进的英国批评者甚至认为艾里有可能与勒维耶串通一气，出卖了亚当斯的计算。这种指控当然是无稽之谈，但艾里与查利斯一度在欧洲大陆及英国本土同时遭到抨击，“猪八戒照镜子，里外不是人”，则是不争的事实。

艾里等人掀起的这场轩然大波不仅极大地伤害了法国人民的感情，而且还严重破坏了英国天文学界自身的和谐，这场风波该如何落幕呢？



注释

[\[1\]](#) 这一细节是艾里于11月13日在皇家天文学会就海王星事件召开的质询会上回顾这一事件时提供的，但史学界对此有一定的争议，因为人们未能查到支持这一说法的文字记录。

[\[2\]](#) 小赫歇耳是否在那次会议上说过那样的话，也同样因为没有找到可以作证的文字记录，而有一定的争议。

[\[3\]](#) 我个人觉得奇怪的是：史料中没有任何有关那段时间小赫歇耳本人从事新行星搜索的记载，以他的家世背景（父亲是天王星的发现者），如果他真的对亚当斯和勒维耶的共同预言有那么大的信心，为何没有亲自搜索新行星呢？

19 握手言和

在所有针对艾里和查利斯的抨击中，有一点无疑击中了要害，那就是在海王星发现之前，他们在一定程度上隐瞒了亚当斯的工作。事先隐瞒，有荣誉时却突然提出，这使得他们的声明在外人——尤其是在法国天文学界——眼里有一种为抢夺荣誉而临时炮制的感觉，成为他们取信于别人的最大障碍。

如果说一开始艾里对亚当斯的工作还只是忽略而非隐瞒，那么在他得知了勒维耶的工作（详见第13章）之后，这样的理由就不大说得通了。很多人认为，艾里和查利斯存在将发现海王星的荣誉留给剑桥的私心，从而有意向同行们隐瞒了亚当斯的工作。这一看法虽从未得到艾里和查利斯的承认，但应该说有一定的合理性⁽¹⁾。艾里本人的说法，则是坚称他自始至终就不曾对亚当斯的工作有足够的重视，即便后来因了解了勒维耶的工作而意识到其结果很可能是正确的，也由于该结果并未正式发表而鲜有提及。但无论出于何种原因，法国方面以此为由全面否认英国方面的声明，甚至认为亚当斯的工作是子虚乌有的骗局，显然是欠冷静的。

如果要盘点一下在发现海王星的过程中英国方面几位当事人的个人过失，那么查利斯显然有着极大的过失。作为一位职业天文学家兼天文台台长，在比对观测数据时如此草率，无论如何是说不过去的。这一点，连他的英国同行们也嗤之以鼻，后来有评论者尖刻地嘲讽道：查利斯是不朽的——他因失败而不朽。

另一方面，艾里虽也饱受抨击，但平心而论，他前前后后的行为倒是都有说得通的理由。比方说亚当斯1845年秋天吃到的几次闭门羹就不能怪艾里，因为亚当斯并未预约。有人也许会对亚当斯第三次登门时艾里正在吃午饭一事感到奇怪，因为当时已是下午四点，但这个古怪的午

饭时间却是艾里医生的要求。而艾里看到亚当斯留下的计算结果后隔了两个星期才回复，则是两个因素的共同结果：一是他的妻子即将生第九个小孩（姜昆和李文华的相声说得好：多子女的坏处就是个人受罪，国家受累）；二是他手下有位职员正好卷入了一桩谋杀丑闻之中。任何人同时碰到这样的家事和公事，恐怕都难免会受到影响。至于他在自己信中所提的天王星轨道径向偏差问题被亚当斯搁置后，不再关注对方，则更是合理的反应。

最后，亚当斯作为这一事件中唯一保持低调的当事人⁽²⁾，虽然得到了英国同行的普遍嘉许，但他没有循正式途径发表自己的计算，无论是因为信心不足，还是为了精益求精，对后来的风波都有直接的负面影响——虽然人们很难拿这一点来批评他。

令人欣慰的是，有关海王星发现的这场轩然大波，在短短几个月之后就在学术界平息了下来。这其中小赫歇耳在遭受法国方面猛烈攻击的情况下坚持斡旋，并用华丽的文字对勒维耶进行安抚，以及艾里在度过了对法国方面公布其私人信件的短暂愤怒期之后采取的克制态度，都起了不小的作用。而英国皇家学会也在这场风波中显示出了非比寻常的气度，将1846年的考普雷奖授予了勒维耶。六十五年前，发现天王星的赫歇耳所获得的第一个崇高荣誉就是考普雷奖（图10），而此时亚当斯尚未获奖，皇家学会就把奖项授予了勒维耶，而且还在获奖理由中称勒维耶的工作是“现代分析应用于牛顿引力理论的最令人自豪的成就之一”，这对勒维耶无疑是极大的安抚⁽³⁾。英国人向来珍视自己的荣誉，这回却将最高荣誉授予了法国方面的竞争者，但英国皇家学会通过这一行为表现出的泱泱气度又何尝不是一种荣誉呢？



图10 考普雷奖章

当然，争论的最终平息还要部分归功于亚当斯的论文。他的论文发表后赢得了一片赞许，很多人（不光是英国人）甚至认为他的方法在数学上比勒维耶的更为优美。法国学术界也最终承认了亚当斯的才华^[4]。1847年6月，亚当斯和勒维耶在英国科学进步协会的一次会议上首度相遇。两人用亲切的交谈开始了他们终生的友谊，也打消了人们对他们会面的担忧。这正是：度尽劫波兄弟在，相逢一笑泯恩仇。

优先权之争的落幕，也终结了勒维耶用自己名字命名海王星的短暂打算。因为这一打算不仅有违行星命名的传统，也与天文学界好不容易达成的亚当斯与勒维耶共享荣誉的共识相违背。

亚当斯与勒维耶这两位当年曾为了请人观测新行星而四处奔走的天体力学高手，后来都亲自担任了天文台的台长：勒维耶于1854年接替去世的阿拉果担任了巴黎天文台的台长，亚当斯则于1861年接替离职的查利斯成为了剑桥天文台的台长。不过他们两位在天文台台长的位置上干得并不出色，亚当斯基本上是把观测事务全都推给了资深助理，勒维耶则不仅同样疏于观测（有人认为他甚至从未在望远镜里看过一眼让他名

垂青史的海王星），而且还因与下属关系恶劣而一度下岗。亚当斯后来两度担任皇家天文学会的主席，在那期间，他亲自向达雷斯特（即与伽勒一起发现海王星的那位学生，他的贡献曾被很多人忽视）和勒维耶颁发过奖项。不过他颁给勒维耶的奖项，我们在后文中将会提到，却是一个乌龙奖项。

海王星事件落幕后，艾里将他手中有关这一事件的信件及其他资料存入了格林威治天文台的档案之中。这些档案被后人称为“海王星档案”（Neptune Files）（图11）。出人意料的是，这些档案在一个半世纪之后又重新掀起了一场风波。

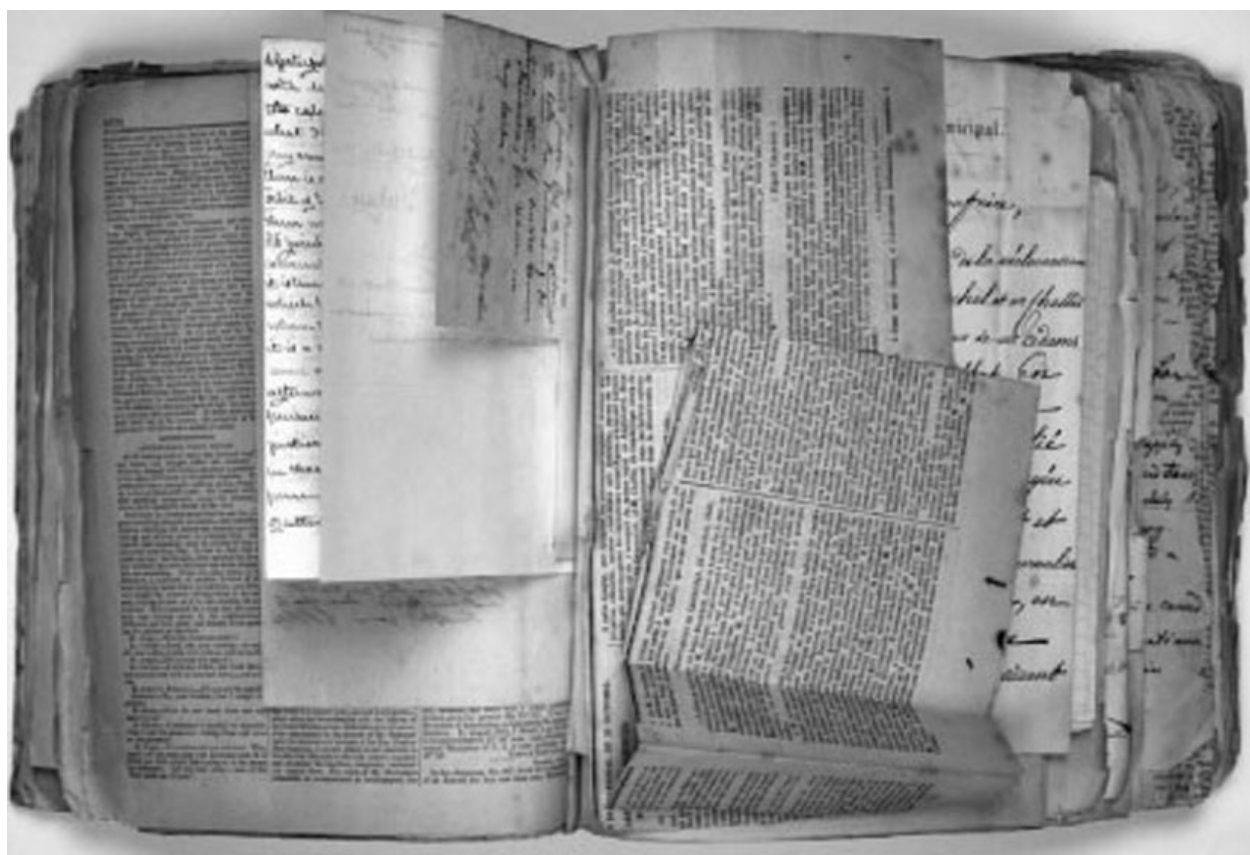


图11 海王星档案

注释

[\(1\)](#) 如我们在第15章中所说，仅凭勒维耶的计算，多数天文学家采取的只是观望态度。因

此知道亚当斯的结果在很大程度上可以算是“剑桥帮”的秘密武器。不过这一看法无法解释艾里为何曾向小赫歇耳等少数同事提及过亚当斯与勒维耶的计算（小赫歇耳虽也曾就读于剑桥，但他并不在剑桥天文台从事观测，应该与艾里设想的剑桥故事没有关系），并明确提出了存在近期内依据这些计算发现海王星的可能性（参阅第13章）。

[\[2\]](#) 后来有人对亚当斯是否真的是一位“timid”（害羞）或“modest”（谦虚）的人提出了异议。但无可否认的是，亚当斯即便在成名之后仍相当低调。他一生谢绝过两次巨大的荣誉：一次是1847年，在小赫歇耳等人的推荐下，维多利亚女王决定授予他爵士头衔，那是牛顿曾经获得过的头衔；另一次则是1881年艾里退休时，他受到推选接替艾里的位置——那是英国天文学界最尊崇的位置。找遍全英国，恐怕也找不出第二位谢绝这两项荣誉的人。

[\[3\]](#) 两年后，即1848年，亚当斯也获得了考普雷奖。

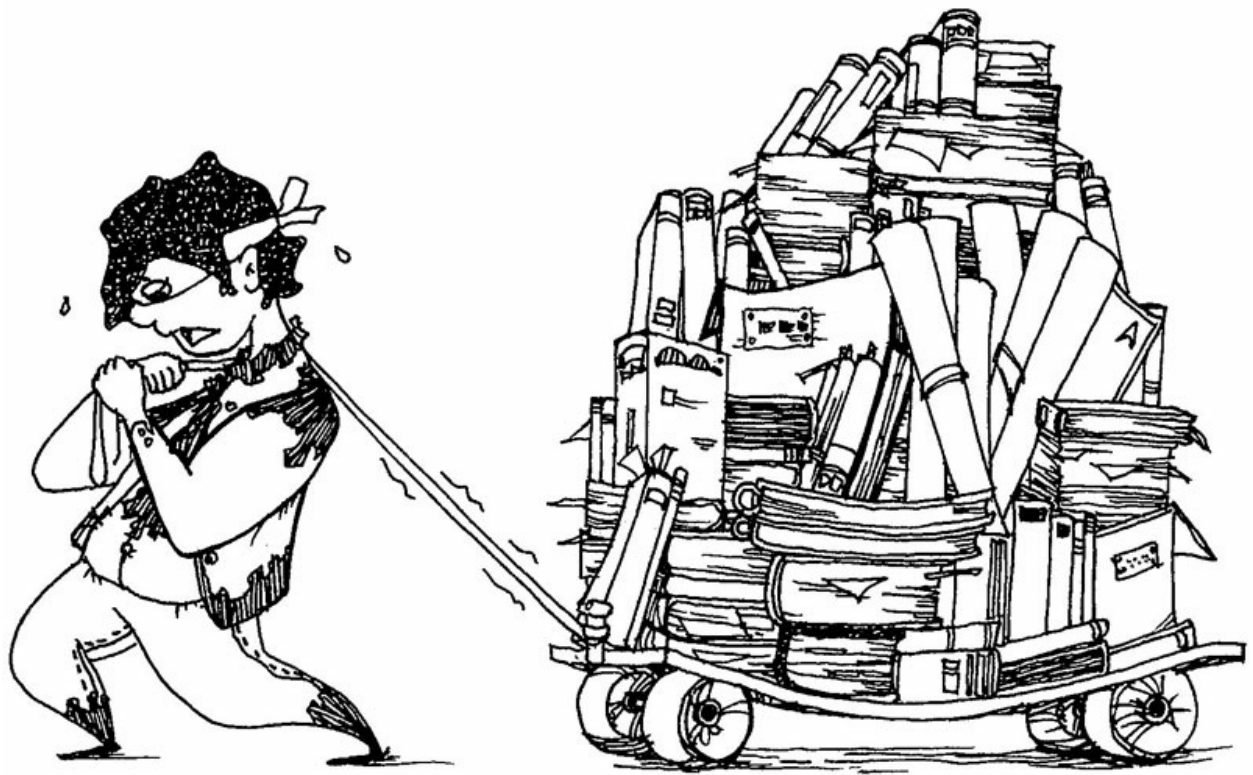
[\[4\]](#) 与勒维耶不同的是，亚当斯的计算细节从未被全部发表，并且他的计算草稿也从未被全部找到。这使得一直有人对亚当斯的工作存有疑虑。不过依据曾经公布过的资料，后人已基本复现了亚当斯的计算方法。

20 秘密档案

海王星的发现在科学界引起了极大的轰动。自那以来，这一发现一直被视为天体力学最辉煌的成就，就像亚当斯与勒维耶的同时代人曾经赞许过的那样。但是，过于夺目的历史光环却也掩盖了这一成就背后的不完美性，以至于后世的很多文章过分渲染了海王星位置与勒维耶的预言相差不到1度这一辉煌之处，却忽略了计算结果中那些与观测不那么相符的地方。

我们在第18章中曾经提到，在海王星被发现之后，亚当斯是第一个利用实际观测数据对其轨道进行计算的天文学家。亚当斯的计算表明，海王星轨道的半长径只有30.05天文单位（现代观测值为30.06天文单位）。稍后，海王星的质量也得到了较为准确的测定，结果表明其质量与天王星几乎相同。将这些结果与亚当斯及勒维耶的计算相比较，不难看到彼此间存在不小的差距。亚当斯的两次计算所采用的天王星轨道半长径分别为38.4和37.3天文单位；勒维耶的两次计算所采用的轨道半长径则分别为38.4和36.2天文单位，均显著大于实际值。而且，除勒维耶的第一次计算采用了圆轨道外，亚当斯和勒维耶所采用的轨道偏心率均在0.1以上，比实际值（约为0.011）大了一个数量级。此外，亚当斯和勒维耶所采用的海王星质量为天王星质量的2~3倍，远远高于实际值。因此，亚当斯和勒维耶的计算无论在天体质量，还是轨道参数上都存在较大的误差。不过幸运的是，对海王星质量的高估，与对其轨道半长径的高估造成的影响在一定程度上得到了抵消，从而大大增加了亚当斯和勒维耶的计算与真实情形的接近程度。即便如此，后来的分析表明，在海王星长达165年的漫长公转周期中，亚当斯和勒维耶的计算只在其中十余年的时间里才是真实轨道的良好近似，而1840-1850年恰好就是这幸运的十年。从这个意义上讲，海王星的发现虽然是一个伟大的天体力学

成就，但它在离计算值如此之近的地方被发现却有一定的偶然性
(1)。



海王星的发现过程是动人心魄的，就连对这一发现过程所做的历史研究也充满了奇峰突起的意外篇章。海王星事件的落幕虽快，却落得并不彻底。一个多世纪以来，一直有人对事件的真相存有疑虑，尤其是对英国方面的说法感到怀疑。终于，这段暗流涌动的历史在相隔一个半世纪后的20世纪90年代末又掀起了一阵新的波澜。

我们在第19章的末尾曾经提到，海王星事件落幕之后，艾里将后来被称为“海王星档案”的一批资料存入了格林威治天文台的档案之中。这些海王星档案在此后一个多世纪的时间里一直处于秘密保存状态，直到二战后的1956年，才随着格林威治天文台的搬迁而重现天日。但颇为离奇的是，这些档案露面后不久就重新失去了踪影。1969年，海王星研究者罗林斯（Dennis Rawlins）在试图查阅海王星档案时，被告知这些档案已不知去向。海王星档案的下落从此成为了一个谜，有人甚至认为这些档案的失踪，乃是英国方面刻意掩盖历史真相的手段。

几十年的时光悄然流逝，海王星档案依旧杳无踪影。1998年，这些档案的昔日藏身之地，有着323年辉煌历史的格林威治天文台因为经费方面的原因而走到了关闭的边缘。世事的变迁似已让这悬案变得越来越没希望了，但就在这“山重水复疑无路”的时候，事情出现了意想不到的转机。1998年10月8日，作为关闭工程的一部分，工人们正准备拆除格林威治天文台的电话线，这时候资深档案管理员珀金斯（Adam Perkins）接到了一个来自遥远的南半球国家智利的电话。电话是从位于智利拉塞里纳（La Serena）的塞罗托洛洛天文台（Cerro Tololo Observatory）打来的。在电话中，珀金斯听到了一个让人几乎不敢相信的消息：失踪了几十年的海王星档案在刚刚去世的恒星天文学家艾根（Olin Eggen）的遗物中被发现了！

原来，有藏书癖好的艾根在20世纪60年代中期利用其在格林威治天文台工作的机会，窃取了包括海王星档案在内的重达百余公斤的档案

[\(2\)](#)。

海王星档案的失而复得很快就在史学界掀起了一场新的波澜。有些人在对那些档案进行研究后，提出了一个惊人的观点，即艾里、查利斯、小赫歇耳、亚当斯等人当年讲述的英国版故事是不真实的，亚当斯在对海王星的预言上不应该享有与勒维耶同等的荣誉。这其中最主要的一位，是一度有过伦敦大学学院荣誉研究员头衔的英国人科勒斯特姆（Nick Kollerstrom）。2001年，科勒斯特姆通过互联网披露了海王星档案的部分内容，并对艾里等人当年的说法提出了全方位的质疑。2003年7月及2004年12月，美国的两份颇具影响力的主流科普杂志《天空和望远镜》（Sky & Telescope）及《科学美国人》（Scientific American）先后刊文介绍了科勒斯特姆的质疑，并且所取标题颇为惊人。《天空和望远镜》的标题为：“秘密档案改写海王星的发现”；《科学美国人》的标题则是：“被盗行星之案”。很多其他媒体也引述或转述了科勒斯特姆的观点，他的正式论文则发表在了2006年3月出版的英国学术季刊《科学史》（History of Science）上。

一时间海王星的发现史似乎重新陷入了重重迷雾之中，英国人真的“盗窃”了海王星，历史真的要被改写吗？

注释

[\(1\)](#) 在后文讲述完冥王星的发现后，我们还会再次谈及这一问题。

[\(2\)](#) 按格林威治天文台后来的说法，档案不是被盗，而是被艾根“借”走了。在艾根去世前，有人曾怀疑是他带走了海王星档案，但他一直予以否认。

21 先入之见

科勒斯特姆对传统海王星发现史的质疑包含了很多方面。从小的方面说，他质疑了传统故事的许多细节，比如亚当斯对艾里的第二和第三次访问（中间相隔一小时）是否真的是在1845年10月21日下午？艾里在他第三次来访时是否真的是在吃午饭？艾里当时到底有没有收到亚当斯的“拜山帖”？艾里是否真的在1846年6月29日的会议期间提及过亚当斯和勒维耶的计算？小赫歇耳是否真的说过发现海王星如同哥伦布从西班牙海岸直接看到美洲这样的话？等等。这些细节从历史研究的严谨性上讲无疑是可以探究的，甚至也可以影响对若干当事人个人过失的大小认定，但很难对事件的整体真实性起到扭转乾坤的作用。

但是从大的方面说，科勒斯特姆的质疑也涉及了一些比较重要的问题。比如我们都知道，亚当斯早在1845年秋天就完成了第一轮计算，并且在访问艾里时留下过一页纸的计算结果。那么，他当时的计算结果究竟是什么呢？传统文献沿用的一直是艾里在海王星发现之后提供的说法，即亚当斯的计算结果与海王星的真实位置只差了 $1^{\circ}44'$ （我们在第12章中所说的“不到两度”指的就是这一说法）。但科勒斯特姆在查阅了一页据称很可能是亚当斯给查利斯的文件，并对比了亚当斯本人的若干笔记后提出，亚当斯当时的计算结果并没有艾里所说的那样精确，而很可能是一个误差达 3° 的结果。科勒斯特姆认为，这样的结果虽然仍是引人注目的，但却不足以引导人们进行有效的搜索。

应该说，科勒斯特姆对这一点的考证是值得重视的，但他的结论却相当突兀，甚至可以说是莫名其妙。 3° 的偏差虽然比 $1^{\circ}44'$ 大了将近一倍，但仍是一个相当小的偏差。若真的有人依据这一结果进行搜索，是完全有可能发现新行星的，因为人们搜索新行星的范围通常都不会定得很小（比如我们在第14章中提到的艾里向查利斯建议的搜索范围就达

30°×10°)。而且更重要的是，我们在前面曾经提到，无论亚当斯还是勒维耶，他们的计算结果与海王星的真实轨道都存在不小的差异。在这种情况下，亚当斯的第一轮计算哪怕真的偏差了3°，也不是什么大不了的问题。甚至哪怕与亚当斯当时计算有关的具体文件已不可考，也不足以改写历史。因为艾里在得知勒维耶的第一轮计算结果后，曾于1846年6月25日在给一位英国同事的信中提到过亚当斯的结果与勒维耶的很接近。当时海王星尚未被发现，我们没有任何理由怀疑艾里在私人信件中所说的那些话。仅此一点，就足以证实亚当斯确实得到过与勒维耶相接近的结果，从而具备与勒维耶分享荣誉的工作基础。

除了对亚当斯第一轮计算的偏差提出质疑外，科勒斯特姆还提到，亚当斯第二轮计算与真实位置的偏差比第一轮更大⁽¹⁾，并且他在1846年9月2日给艾里的信中曾对自己的预测作过幅度高达23°的错误变动。科勒斯特姆据此认为，亚当斯既没有稳定的计算结果，也不具备对自己计算的基本自信。应该说，与勒维耶相比，亚当斯在自信心上的确显得比较欠缺。不过我们对他们工作的评价，首要的依据是他们的计算方法是否正确，以及他们的计算结果能否对实际观测起到引导作用。受当时的计算能力（尤其是数值计算能力）所限，他们两人的计算误差都是比较大的，勒维耶的计算误差达10°左右，亚当斯的有可能更高。在这样的误差下，第二轮计算的实际偏差是变大还是变小，并不能有效地衡量他们计算方法的优劣，甚至也不能作为判断他们计算误差的充分依据。至于亚当斯对自己预言所作的巨幅变更，据分析很可能是因为将瑞士天文学家瓦特曼1836年公布的一组错误数据视为了新行星的观测位置（因为瓦特曼在公布数据时曾宣称那是他观测到的新行星），与他计算方法的正确与否无关。而且那次巨幅变更只是一次孤立的预言，与他的两轮系统计算并无实质关联。退一步说，即便勒维耶的计算的确比亚当斯更为精确，甚至精确很多，但从上文提到的艾里给同事的信件，以及艾里因两人的预测相近而催促查利斯进行观测来看，亚当斯的结果也仍

足以对实际观测起到引导作用。因此，这方面的质疑同样不足以改写历史。

如果说上面那些质疑还只是单纯的技术性质疑，所涉及的只是亚当斯计算的技术水准，那么科勒斯特姆的另一类质疑，则把锋芒指向了艾里等人的诚信。在这类质疑中，他通过对艾里、查利斯等人的文章及信件（尤其是信件）中各种细节乃至语气的辨析，指出他们有可能在有关这一事件的若干叙述中撒了谎。这种辨析在当年优先权之争最炽热的时候，勒维耶、阿拉果等法国天文学家也曾用过（参阅第18章），只不过科勒斯特姆做得更为系统，也更加详尽。

不过，这些辨析究竟有多大说服力，是值得商榷的，而凭借那些辨析对这么重大的历史事件进行翻案，则更值得怀疑。因为我们都知道，信件的内容常常会因收信人的不同而有不同的侧重点。比如在试图安抚法国同行的时候，艾里就会有意突出后者的贡献，少提或不提亚当斯，以免产生副作用。而信件的语气则不仅与收信人有关，还与写信人的心情有关，不同的语气体现的有可能只是心情的差异，甚至相互间的矛盾也可能只是记忆的差错或笔误。信件不是论文，是不会有编辑来替写信人修改笔误的。事实上，科勒斯特姆能从艾里等人的信件中看出那么多的“问题”，与其说是表明艾里等人有可能撒了谎，不如说是恰恰说明他们并未撒谎。因为那些信件大都是海王星事件发生之后所写的，以艾里等人的智力，倘若有意要编造故事，又岂会在那些后期信件中留下如此多的破绽？那些“破绽”出现在普通信件中是可以理解的，但作为三个著名学者合谋故事的一部分，却是根本不应该出现的。更何况，如果艾里等人真的撒了谎，艾里又为何要留下海王星档案来让后人追查真相？再说小赫歇耳和查利斯早在10月3日就各自发表文章提及了亚当斯的贡献（参阅第18章），当时艾里尚在欧洲大陆旅行。他们若要编故事，又怎敢在艾里这么重要的知情人返回英国相互协调之前就贸然行事？

总体来说，科勒斯特姆对海王星事件的研究带有较强的先入之见，

即首先认定失踪档案隐藏着重大问题，然后去寻找证据。这种“史从论出”的“阴谋论”心态是史学研究的大忌，带着这种心态研究史料，很容易把一些并无充分说服力的细节视为铁证，赋予它们不应有的重要性，就像中国寓言故事“疑人偷斧”所隐喻的那样。而且一旦有了先入之见，常常会有意无意地忽略或回避对自己观点不利的东西，千方百计地穿凿附会自己早已设定的结论，从而丧失客观公正的立场⁽²⁾。艾里、查利斯及小赫歇耳都是有名望的天文学家，作为当时英国天文界的重要人物，他们当然很看重英国天文界的整体荣誉，但认为他们会在如此重大的学术事件中编造谎言，是令人难以置信的。因为这种谎言一旦败露，将对英国的学术声誉带来重大灾难。更何况，除小赫歇耳外，艾里和查利斯都在海王星事件中遭受了巨大的个人名誉损失（若亚当斯并未独立推算出海王星的位置，或他的工作质量与勒维耶不可相提并论，那么后人加诸于艾里和查利斯的恶评无疑会少得多）。科勒斯特姆提出的“证据”显然远不足以解释这几位功绩卓著的天文学家为何要用自己宝贵的名誉，来进行一场吉凶未卜的豪赌，并且赌得如此粗心，甚至还特意保留了“罪证”。

海王星档案的失而复得有助于史学界更精确地还原海王星发现过程中的若干细节，但起码就目前看到的资料和分析而言，它完全不足以改写历史。海王星的发现是科学界的一个伟大成就，亚当斯和勒维耶各自独立地计算出了海王星的位置，而伽勒及达雷斯则一同发现了这颗新行星。

注释

⁽¹⁾ 科勒斯特姆在这点上自相矛盾的。亚当斯第二轮计算的偏差为 $2^{\circ}30'$ ，而他第一轮计算的偏差——按照科勒斯特姆自己的考证——则是 3° 。因此，所谓第二轮计算的偏差比第一轮更大的说法是与他自己的考证相矛盾的。这一矛盾说明科勒斯特姆重新将艾里所说的 $1^{\circ}44'$ 作为亚当斯第一轮计算的偏差（作为对比，勒维耶第二轮计算的偏差由第一轮 $-2^{\circ}21'$ 缩小

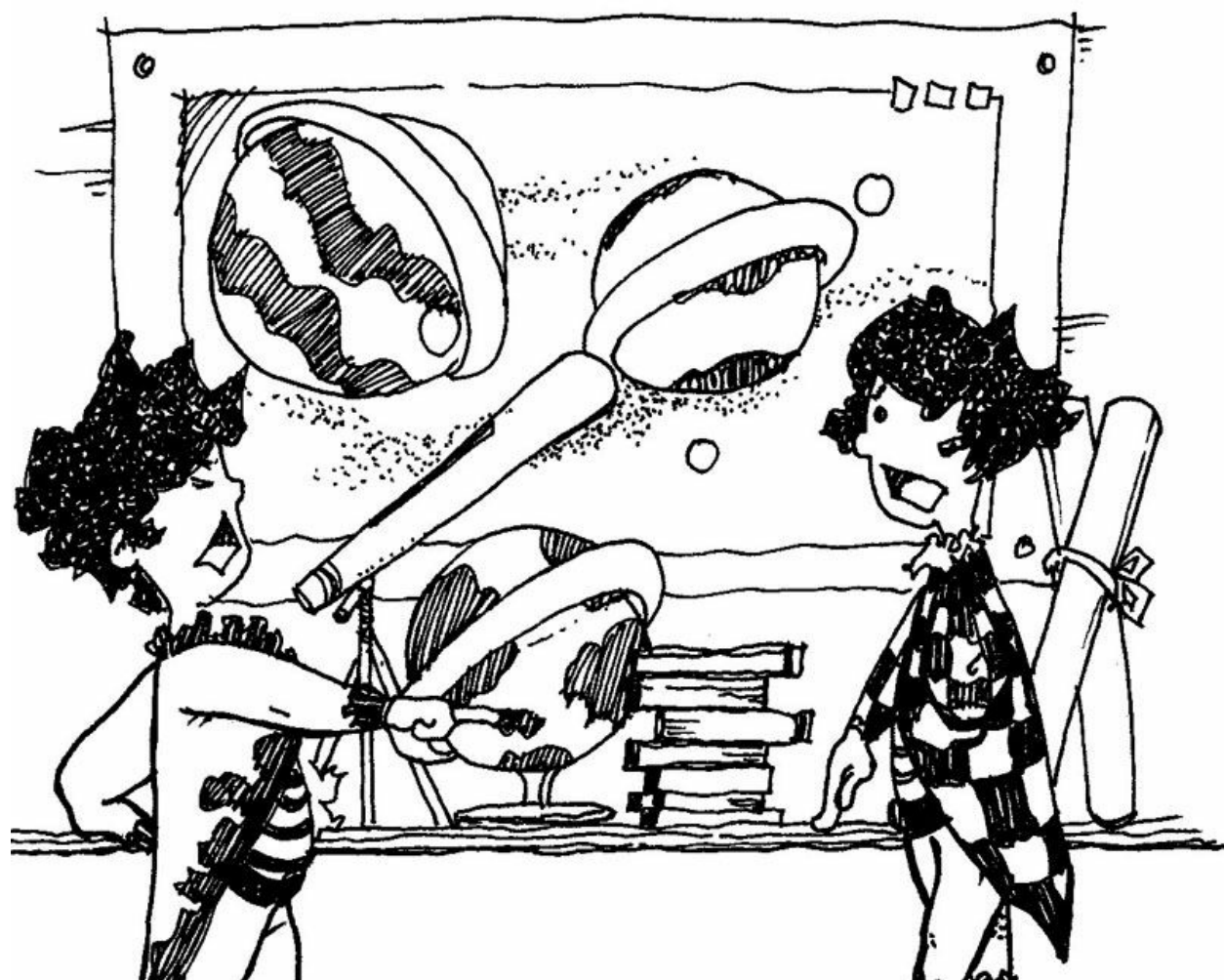
为 $-0^{\circ}58'$)。这种视自己需要而随意选用彼此矛盾的数据的做法显然是有失严谨的。另外值得一提的是，科勒斯特姆认为亚当斯和勒维耶的第二轮计算之间的相互差异有 3.5° ，而非一些早期文献所说的不到 1° 。

[\[2\]](#)这一点也正是科勒斯特姆的致命弱点，他对海王星发现史的质疑虽曾被一些主流科普杂志、学术刊物及媒体所引述，但他的历史“研究”有着浓厚的伪历史及阴谋论色彩。除质疑海王星的发现史外，他还质疑纳粹大屠杀的真实性，是所谓的“大屠杀否认者”(holocaust denier)之一，并因此于2008年4月被伦敦大学学院撤销了一切学术头衔。

22 火神疑踪

海王星的发现极大地刺激了天文学家和数学家的兴趣。原本属于观测天文学家专利的新行星，居然可以用纸和笔来发现，这实在太吸引人了。一时间用数学方法寻找新行星成为了时尚。天文学家们兵分两路展开了行动，一路沿袭了向外扩张的历史传统，到海王星轨道之外去寻找惊喜；另一路则独辟蹊径，将目光投向了水星轨道的内侧。这后一路天文学家的领军人物不是别人，正是赫赫有名的勒维耶。在发现海王星的荣誉出人意料地被亚当斯分走一半后，勒维耶决定寻找一个新的猎物——一个自己可以独享的猎物。当时多数天文学家认为在海王星之外发现新行星的机会更大，但勒维耶却认为在距海王星的发现如此之近，从而对海王星轨道的了解还不充分的情况下，用数学手段寻找新行星尚为时过早。因此，虽然他也相信海王星之外存在新的行星，但却首先选择将水星轨道以内作为自己的新战场。

勒维耶之所以选择水星轨道以内作为新战场，还有一个很重要的原因，那就是水星的轨道也存在着反常。经过长期精密的观测，天文学家们早就发现水星的椭圆轨道在背景星空中存在缓慢的整体转动，这种转动被称为水星的近日点进动。观测表明，水星的这种近日点进动平均每年约为56〔角〕秒。但另一方面，考虑了由地球自转轴进动造成的表观效应及已知行星的影响后，理论计算给出的进动值却只有每年55.57〔角〕秒⁽¹⁾，两者相差0.43〔角〕秒。天文学家们知道水星轨道的这一细微反常已有时日，勒维耶本人早在当年对各大行星做地毯式研究（参阅第13章）时，就曾对水星轨道进行过详尽考察。海王星的发现无疑赋予了这一反常一个全新的意义。在勒维耶看来，这个虽然微小，但确凿无疑的轨道反常，是水星轨道之内存在未知天体的明显证据。



那么这未知天体会是个什么样的天体呢？勒维耶认为有两种可能性：一种是单一行星，另一种则是小行星带。也许是由于水星近日点的反常进动与当年的天王星出轨相比显得更为规则，或者是受当时正在发现中的小行星带的启示，勒维耶比较倾向于后一种可能性，即在水星轨道之内存在一个小行星带。1859年9月，他在一篇文章中正式预言在距太阳0.3天文单位处存在一个未被发现的小行星带。

正所谓：说曹操，曹操到。勒维耶的预言提出后不久，一位名叫莱沙鲍特（Edmond Lescarbault）的法国医生兼业余天文学家就给他写来了一封信，声称自己曾于1859年3月26日发现过一个穿过太阳表面的天体。这封来信让勒维耶很是兴奋，他立即对这位业余天文学家进行了“家访”。在确信此人值得信赖后，勒维耶依据他所得到的数据对这一天体的参数进行了计算，结果表明其轨道半径为0.147天文单位，质量约为水星质量的百分之六。这个天体很快就被取名为火神星（Vulcan，罗马神话中的火神及希腊神话中的工匠之神，美神维纳斯的丈夫）。1860年初，勒维耶向法国科学院报告了发现火神星的消息。尽管自首次“发现”以来，包括莱沙鲍特本人在内的所有人都不曾再有机会一睹火神星的芳容，但法国科学院基于对勒维耶的无比信任，还是很痛快地将拿破仑设立的法国最高勋章——军团勋章（Légion d'honneur）授予了莱沙鲍特，从而上演了该院历史上最大的乌龙颁奖事件之一。

虽然火神星的轨道半径远小于勒维耶预言的0.3天文单位，其引力作用也远不足以解释水星近日点的反常进动，但勒维耶一生都对它的存在深信不疑。受他的巨大声望影响，一些天文学家在此后近20年的时间里锲而不舍地找寻着火神星的情影，其中包括在浩如烟海的文献中搜寻可能存在的历史记录。1876年，在亚当斯担任主席期间，英国皇家天文学会也步法国科学院的后尘，很乌龙地在火神星的存在尚未得到确认的情况下，就将一枚金奖授予了勒维耶，以表彰他为解决水星近日点反常

进动问题所做的贡献。

但这一切的热情都没能感动火神星，这颗神秘的“行星”再也不曾露面过，所有曾被当作火神星的历史记录（主要集中在1819-1837年间）也都被一一判定为是太阳黑子而非天体。1877年9月23日，火神星的最大支持者勒维耶离开了人世，这一天距海王星的发现正好相隔31年，但火神星的命运仍悬而未决。

火神星之所以能在那么长的时间内杳无踪迹，却仍让那么多的天文学家牵肠挂肚，除了依靠勒维耶的“魅力值”外，一个很重要的原因是它离太阳太近，太容易湮没在太阳的光芒之中，从而即便长时间观测不到，也无法说明它不存在。

但丑媳妇终究是要见公婆的。1878年7月29日，天文学家们迎来了一个搜寻火神星的绝佳机会：日全食。当太阳的光芒不再夺目时，火神星还如何遁迹？那一天，大批天文学家在可以观测日全食的美国怀俄明州的一个小镇上架起了望远镜，等待火神星之谜的水落石出。

但出人意料的是，那天的观测没能对火神星的命运作出宣判，却充分证实了心理学的巨大威力。那一天，不相信火神星的天文学家们全都没有观测到火神星，从而更坚信了火神星的子虚乌有⁽²⁾。但相信火神星的职业天文学家沃森（James Watson）及业余天文学家斯威福特（Lewis Swift）却都声称观测到了火神星，斯威福特甚至声称自己观测到了两个水内天体。虽然这两人宣称的天体位置彼此之间以及与勒维耶的预言之间全都不同（从而无法相互印证），而且很快就有天文学家通过他们记录的天体位置指出他们很可能将已知天体误当成了火神星，但这两位老兄爱火神星没商量，一口咬定自己观测到的就是火神星。

在那之后又过了十几年，人们在勒维耶有关火神星轨道的计算中发现了错误。不仅如此，进一步的分析表明，火神星的存在与其他内行星——尤其是金星——的运动并不相容。自那以后，火神星的追随者基本上销声匿迹了。

最终为火神星的疑踪画下完美句号的是物理学家爱因斯坦（Albert Einstein）。1915年，他在刚刚完成的广义相对论的基础上，完美地解释了水星近日点的反常进动，从而彻底铲除了火神星赖以存在的理论土壤^{[\(3\)](#)}。

注释

^{[\(1\)](#)}这其中由地球自转轴进动造成的表观效应约为每年50.256〔角〕秒，由已知行星的引力作用产生的进动约为每年5.314〔角〕秒。

^{[\(2\)](#)}从理论上讲，在日全食期间没有观测到火神星并不意味着火神星不存在，因为它有可能恰好也和太阳一起被遮盖。不过这种情况发生的概率较小（感兴趣的读者可以估计一下这一概率的大小）。

^{[\(3\)](#)}即便如此，仍有个别天文学家在水星轨道以内寻找新天体。不过这类天体的线度上限已被压缩到了60千米，至多只能是小行星。

23 无中生有

寻找火神星的天文学家们已全军尽墨，但在海王星以外寻找新行星的天文学家们却还处在忙碌之中，他们的战场完全是另一番景象。

我们知道，海王星之所以能在笔尖上被发现，是因为天王星存在出轨现象，而勒维耶之所以寻找火神星，是因为水星也存在出轨现象，虽然那种被称为水星近日点反常进动的出轨现象具有高度的规则性，从而与天王星的出轨完全不同。那么，寻找海王星以外的行星（以下简称海外行星），尤其是通过计算手段寻找那样的行星，它的依据又在哪里呢？很遗憾地说，只存在于天文学家们那些“骚动的心”里。

自从海王星被发现之后，天王星的出轨之谜基本得到了解释，剩余的偏差已微乎其微。但如何看待这细微的剩余偏差，却有很大的讲究。我们知道，有关行星轨道的任何观测及计算都是有误差的，因此计算所得的轨道与观测数据绝不可能完全相符。一般来说，只要两者的偏差足够小，小于观测及计算本身所具有的误差，这种偏差就算是正常的，并且往往是随机的。天王星的出轨与水星近日点的反常进动之所以引人注目，是因为它们都远远超过了观测及计算的误差。但是，海王星被发现之后，天王星的剩余“出轨”实际上已经处在观测及计算误差许可的范围之内，没有进一步引申的余地了。不幸的是，发现海王星的成就实在太令人心醉，以至于此前一直追求观测与计算的一致，并愿为之奋斗终生的一些天文学家，现在反而由衷地期盼起观测与计算的不一致来。因为唯有那样，才有重演海王星发现史的可能。正是在这种满心的期待乃至虔诚的祈祷之中，天文学家们开始在鸡蛋里挑骨头，他们的目光变得多疑，他们不仅“发现”天王星仍在出轨，而且怀疑海王星也不规矩。



1848年，距海王星的发现仅仅过了两年，法国天文学家巴比涅特（Jacques Babinet）就预言了一颗海外行星。他提出的海外行星的轨道半长径约为47~48天文单位，质量约为地球质量的11.6倍。他的计算依据是海王星的实际轨道与勒维耶所预言的轨道之间的差别。显然，这是一种完全错误的计算逻辑。因为勒维耶所预言的轨道只是依据天王星出轨现象所作的推测，而且在推测时还对轨道参数（比如半长径）做过带有一定任意性的猜测，从而根本就不是标准的海王星轨道计算。（请读者想一想，标准的海王星轨道计算应该是怎样的？）用那样的轨道来研究海王星的出轨，套用著名物理学家泡利（Wolfgang Pauli）的话说，那是“连错误都不如”（not even wrong）。

理论天文学家们的心情固然急切，观测天文学家们的动作也不含糊。1851年，距海王星的发现仅仅过了四年多，英国天文学家辛德（我们在第18章中提到过此人，他是海王星被发现后第一位观测海王星的英国人）宣布自己从美国天文学家弗格森（James Ferguson）的一份观测报告中，发现了一颗轨道半长径为137天文单位的海外行星。但是，无论辛德、弗格森还是其他人，都没能再次捕捉到那颗神秘的“海外行星”，它的谜底直到28年后才揭晓，原来那是弗格森的一次错误的观测记录^{[\(1\)](#)}。

这些早期的谬误并未阻止更多的天文学家对海外行星作出预言。从19世纪中叶到20世纪初的50年间，欧洲和美国的天文学家们轮番向海外行星发起了冲击，并取得了如下“战果”：

- 托德（David Todd）预言了一颗海外行星，轨道半长径为52天文单位。

- 弗莱马力奥（Camille Flammarion）预言了一颗海外行星，轨道半长径为45天文单位。

- 福布斯（George Forbes）预言了两颗海外行星，轨道半长径分别

为100和300天文单位。

·劳（Hans-Emil Lau）预言了两颗海外行星，轨道半长径分别为46.6和70.7天文单位。

·达利特（Gabriel Dallet）预言了一颗海外行星，轨道半长径为47天文单位。

·格里戈尔（Theodore Grigull）预言了一颗海外行星，轨道半长径为50.6天文单位。

·杜林冈德斯（Vicomte du Ligondes）预言了一颗海外行星，轨道半长径为50天文单位。

·西伊（Thomas See）预言了三颗海外行星，轨道半长径分别为42.25、56和72天文单位。

·伽诺夫斯基（Alexander Garnowsky）预言了四颗海外行星，但没有提供具体数据。

一时间外太阳系几乎变成了计算天文学的练兵场。在上述计算中，除天王星和海王星的轨道数据外，有些计算（比如弗莱马力奥和福布斯的计算）还利用了某些彗星的轨道数据。但与亚当斯和勒维耶对海王星的预言截然不同的是，天文学家们对海外行星的预言无论在数量、质量、轨道半长径，还是具体方位上都是五花八门。如果一定要从那些预言中找出一些共同之处，那就是“三不”：即全都不具有可靠的理论基础，全都不曾得到观测的支持，以及全都不靠谱。

为什么亚当斯与勒维耶预言的海王星参数彼此相近，而人们对海外行星的预言却如此五花八门呢？这个并不深奥的问题终于引起了一位法国天文学家的注意。此人名叫盖洛特（Jean Baptiste Gaillot），他对天王星和海王星的轨道进行了仔细分析，得出了一个直到今天依然正确的结论：那就是在海王星被发现之后，天王星和海王星轨道的观测数据与理论计算在误差许可的范围之内已经完全相符。换句话说，天王星的出轨问题已经因为海王星的存在而得到了完全的解释，在误差许可的范围

之内，根本就不存在所谓天王星的剩余出轨或海王星的出轨问题。

盖洛特的分析很好地解释了为什么天文学家们有关海外行星的预言如此五花八门，却无一中的。记得很多年前笔者曾经读到过一则小故事，说有三位绘画爱好者去拜访一位名画家。在画家的画室里他们看到了一幅刚刚完成的山水画，那画很漂亮，但令人不解的是，在画的角落上却有一团朦胧的墨迹。这三人深信那团墨迹必有深意，于是便对其含义作出了五花八门的猜测。后来还是画家本人为他们揭开了谜底：原来那墨迹是画家的孙子不小心弄上去的。在天文学家们预言海外行星的故事中，观测与计算的误差仿佛是那团墨迹，它本无深意，醉心于海王星发现史的天文学家们却偏偏要无中生有地为它寻求解释，从而有了那些五花八门的预言。

分析是硬道理，事实更是硬道理，在亲眼目睹了那么多的失败预言后，多数天文学家接受了盖洛特的结论，认为像预言海王星那样从理论上预言海外行星，起码在当时的条件下是不可能的。不过预言海外行星的努力并未就此而终止，因为有两位美国天文学家偏偏不信这个邪，他们誓要将对海外行星的预言进行到底。

注释

[\(1\)](#) 这一错误是美国天文学家彼得斯（Christian Peters）所发现的。

24 歧途苦旅



美国天文学家皮克林（1858-1938）

这两位歧途上奋勇前进的美国天文学家对新行星的预言风格恰好走了两个极端。一位犹如天女散花，四面出击；另一位则谨记传统方法，抱元守一。皮克林（William Pickering）是那位喜欢天女散花的预言者。此人出生在美国的波士顿，这是世界名校哈佛大学与麻省理工学院的所在地，有着厚重的学术沉淀。皮克林有位兄弟担任过哈佛学院天文台（Harvard College Observatory）的台长⁽¹⁾，而他本人在天文领域也小有成就，曾于1899年发现了土星的一颗卫星，不过他也热衷于研究一些后来被证实为子虚乌有的东西，比如月球上的昆虫和植被。总体来

说，皮克林的工作风格不够严谨，这在很大程度上影响了他的学术成就，他一生有过的最高学术职位只是助理教授。皮克林晚年花了大约20年的时间研究海外行星，他在这方面的研究很好地示范了他的马虎风格。他虽然是一个人在战斗，但提出的海外行星数量之多，更改之频，信誉之低，以及参数之千差万别，全都堪称奇观。自1908年提出第一个预言以来，他先后预言过的海外行星共有7颗之多，且四度更改预言，他用英文字母对自己的行星进行了编号。为了对他的“战果”有一个大致了解，我们将他的预言罗列一下（其中行星U的轨道虽在海王星以内，却也是为了解释天王星和海王星的“出轨”而提出的；带撇的行星则是相应的不带撇行星的“补丁加强版”）：

- 行星O（1908年）：轨道半长径51.9天文单位，质量为地球质量的2倍。

- 行星P（1911年）：轨道半长径123天文单位。

- 行星Q（1911年）：轨道半长径875天文单位，质量为地球质量的20000倍。

- 行星R（1911年）：轨道半长径6250天文单位，质量为地球质量的10000倍。

- 行星O'（1919年）：轨道半长径55.1天文单位，质量为地球质量的2倍。

- 行星O''（1928年）：轨道半长径55.1天文单位，质量为地球质量的0.75倍。

- 行星P'（1928年）：轨道半长径67.7天文单位，质量为地球质量的20倍。

- 行星S（1931年）：轨道半长径48.3天文单位，质量为地球质量的5倍。

- 行星T（1931年）：轨道半长径32.8天文单位。

- 行星P''（1931年）：轨道半长径75.5天文单位，质量为地球质量的

50倍。

·行星U（1932年）：轨道半长径5.79天文单位，质量为地球质量的0.045倍。



除孜孜不倦地从事计算外，皮克林还投入了大量的时间亲自搜索这些新行星。可惜他预言的行星虽多，在观测上却一无所获。1908年，在他完成了自己的第一个预言——对行星O的预言——后，他向一位名叫罗威尔（Percival Lowell）的美国天文学家请求了观测方面的协助。这位罗威尔是他的波士顿老乡，而且很巧的是，罗威尔也有一个兄弟在哈佛任职，且职位更牛，曾任哈佛校长^[2]。与皮克林研究月球上的昆虫和植被相类似，罗威尔也热衷于研究一些后来被证实为子虚乌有的东西，比如火星人和火星运河。罗威尔对天文学的主要贡献是，出资在亚里桑那州（Arizona）的一片海拔两千多米的荒凉高原上建立了著名的罗威尔天文台（Lowell Observatory）。这是美国最古老的天文台之一，也是全世界最早建立的远离都市地区的永久天文台之一。这一天文台早期的一个主要使命就是观测火星运河。



美国天文学家罗威尔 （1855-1916）

皮克林之所以请求罗威尔提供协助，除两人是同乡兼同行外，还有一个原因，那就是皮克林曾在罗威尔天文台的兴建过程中向罗威尔提供

过帮助。按说有这么多层的“亲密”关系，罗威尔是没有理由不鼎力相助的。可惜皮克林却有一事不知，那就是罗威尔正是那另一位“不信邪”的美国天文学家，他当时也在从事新行星的搜寻工作，而且已经进行了三年。有亚当斯与勒维耶的海王星之争作前车之鉴，罗威尔对自己在这方面的努力进行了严格的保密，甚至在天文台内部的通信中都绝口不提新行星一词。接到皮克林的请求后，罗威尔暗自心惊。他一方面不动声色地予以婉拒，另一方面则加紧了自己的努力，将精力从火星运河上收了回来，集中到对新行星的研究上来。不过当他看到皮克林的粗糙计算后，立刻就放了心，看来并不是什么人都有能力从事这方面的工作的。自那以后，罗威尔不再避讳提及新行星，他将新行星称为行星X。

罗威尔寻找新行星的努力最初侧重的是观测，可惜一连五年颗粒无收。自1910年起，他决定对新行星的轨道进行计算，以便为观测提供引导。罗威尔的数学功底远在皮克林之上，与后者的漫天撒网不同，罗威尔对新行星的计算具有很好的单一性（即相信所有的剩余“出轨”现象都是由单一海外行星造成的）。与亚当斯和勒维耶一样，他首先对新行星的轨道半长径作出了一个在他看来较为合理的假设，然后利用天王星和海王星的“出轨”数据来推算其他参数。在具体的计算上他采用了勒维耶的方法（因为勒维耶发表了完整的计算方法，而亚当斯只发表了一个概述）。

那么新行星的轨道半长径应该选多大呢？罗威尔进行了独特的分析。由于海王星的发现明显破坏了提丢斯一波德定则，因此在寻找海外行星时人们已不再参考这一定则。为此，罗威尔提出了一个新的经验规律，那就是每颗行星与前一颗行星的轨道周期之比都很接近于一个简单分数，比如海王星与天王星的轨道周期之比约为2：1，土星与木星的轨道周期之比约为5：2。在此基础上，他提出一个假设，即行星X与海王星的轨道周期之比是2：1。由开普勒第三定律可知（请读者自行验证），这意味着行星X的轨道半长径约为47.5天文单位。应该说，罗威

尔的这个猜测有其高明之处，因为某些行星（或卫星）的轨道之间存在着所谓的轨道共振现象，它们的周期之比的确非常接近简单分数。不过轨道共振并非普遍现象⁽³⁾，即便出现轨道共振，也没有理由认为行星X与海王星的轨道周期之比就一定是2：1⁽⁴⁾。罗威尔自己或许也意识到了这一点，他后来还尝试过两个不同的轨道半长径：43.0和44.7天文单位。1912年，劳累过度的罗威尔病倒了几个月，但借助四位数学助手的协助，他终于在1913-1914年间完成了初步计算，他给出的行星X的质量为地球质量的6.6倍。

在进行理论计算的同时，罗威尔也没有放弃观测搜寻。他将自己一生的最后岁月全都投入到了搜寻新行星的不懈努力之中。可惜的是，他——以及皮克林——的所有努力与以前那些失败的预言并无实质差别。如果把他们投入巨大心力所做的计算比喻为大厦，那么所有那些大厦——无论多么华美——全都是建立在流沙之上的。随着时间的推移，罗威尔的努力越来越被人们所忽视。1915年初，他在美国艺术与科学学院（American Academy of Arts and Science）所作的一个有关海外行星搜索的报告受到了学术界与公众的双重冷遇，他的文章甚至被科学院拒收。自那以后，罗威尔对新行星的热情一落千丈，而他的生命之路也在不久之后走到了尽头。

1916年，罗威尔带着未能找到行星X的遗憾离开了人世。在他一生的最后五年里，罗威尔天文台积累了多达1000张的照相记录，在那些记录中包含了515颗小行星，700颗变星⁽⁵⁾，以及——他万万不曾想到的——新行星的两次影像⁽⁶⁾！这真是：有缘千里来相会，无缘对面不相逢。

注释

⁽¹⁾ 皮克林的这位兄弟名叫爱德华（Edward Pickering），于1877-1919年间任哈佛学院天文台的台长。原子光谱中的皮克林线系（Pickering series）就是以皮克林的这位兄弟的名字命名

的，他并且还是分光双星（spectroscopic binary）的发现者。

[\[2\]](#) 罗威尔的这位兄弟名叫阿伯特（Abbott Lowell），于1909-1933年间任哈佛大学校长。

[\[3\]](#) 由于太阳系相邻行星（小行星带也算在内）自外而内的轨道周期之比都在1~3之间，即便不存在轨道共振，它们接近于简单分数的概率也不小。感兴趣的读者可以算一下，任意一个1~3之间的实数与一个简单分数（比如分子分母都不超过5）接近到8%（这是罗威尔的猜测对已知行星的最大误差）以内的概率有多大。

[\[4\]](#) 如果把后来发现的冥王星视为行星X的话，它与海王星则的确存在轨道共振现象，只不过它们的周期比是3：2而不是2：1。

[\[5\]](#) 变星通常显示为亮度变化的天体，与移动天体明显不同。但有些变星在亮度变小后会因为比相片所能记录的最暗淡的天体还要暗，而从相片中消失，这样的变星在闪视比较时很像是一颗移出（或移入）相片范围的移动天体。

[\[6\]](#) 那是1915年4月7日由他的助手比尔（Thomas Bill）所做的观测记录，那时罗威尔自己已不再从事观测。

25 农家少年

罗威尔虽然去世了，但他为自己的未竟事业留下了一份最宝贵的遗产，那就是罗威尔天文台。他还在遗嘱中留出了超过一百万美元作为天文台的运作经费，这在当时是一个巨大的数目。可惜的是，第一次世界大战的爆发彻底终止了像寻找新行星那样的“小资”活动。更糟糕的是，罗威尔的遗孀因不满财产分配而发起了一场诉讼官司，这场官司不仅阻碍了天文台的运作，而且耗去了罗威尔留给天文台的那笔经费的很大一部分。经历了这些波劫的天文台直到1927年才重回正轨，可经费却已变得拮据。这时候，罗威尔那位担任哈佛校长的哥哥伸出了援助之手，向天文台捐赠了一万美元。在此基础上，天文台开始装备一台口径13英寸的照相反射望远镜（图12）。



图12 发现冥王星所用的望远镜

不过世事变迁对罗威尔天文台的影响不仅体现在财务上，也涉及了

学术。当时罗威尔的多数工作（比如对火星运河的观测）已被天文学界判定为是毫无价值的，而大半个世纪以来有关新行星的天女散花般的“预言”也早已信誉扫地。天文台是否还要继承“罗威尔道路”呢？罗威尔生前从事的寻找新行星的工作是否还要继续呢？这是罗威尔天文台面临的一个新的十字路口。在这个路口上，天文台的资深天文学家们大都作出了与当年那些错过了海王星的天文学家们一样的选择，即用其他任务填满自己的工作日程，不再抽时间从事新行星的搜索。对于一般的天文台来说，这应该就是新行星故事的终结了。不过罗威尔天文台终究不是一般的天文台，它并未完全忘记创始人罗威尔的心愿。虽然不可能再以新行星搜索为工作重心，但它当时的托管人——罗威尔的外甥普特南（Roger Putnam）——决定招募一名观测助理来从事新行星的搜索。

说来也巧，恰好就在这时，一封来自堪萨斯州（Kansas）的求职信寄到了天文台，求职者是一位22岁的农家少年。

这位少年名叫汤博（Clyde Tombaugh），1906年2月4日出生在伊利诺伊州（Illinois），16岁时随父母迁居到堪萨斯州。受他叔叔的影响，汤博从小喜爱天文。由于家境贫寒，加上父母生育了6个孩子，汤博中学毕业后只能辍学在家。他白天帮家里干农活，晚上则沉醉于观测无穷无尽的星空。由于没钱购买合适的望远镜，汤博用废弃的船舱玻璃、木板及农机零件，自己动手制作了口径为7英寸和9英寸的望远镜。



美国天文学家汤博 （1906-1997）

如果不是1928年的一场突如其来的冰雹，汤博的一生也许就这样静静地农庄里度过了。那一年，汤博家的庄稼长势极好，却在收获季节来临之前毁于冰雹。这场变故让汤博觉得应该找一个更可靠的职业来资助家里。于是他向当时自己知道的唯一一个天文台——罗威尔天文台——发去了求职信，并在信中附上了自己的一些笔记和图片。



一位务农在家且只有中学学历的小伙子能引起罗威尔天文台的注意吗？很幸运，答案是肯定的。汤博在求职信中所附的笔记和图片给罗威尔天文台的台长斯莱弗（Vesto Slipher）留下了很好的印象。他制作望远镜的手艺也正是罗威尔天文台所需要的，因为天文台的13英寸照相反射望远镜当时正在装配之中。甚至连他的务农经历对斯莱弗来说也显得很亲切，因为斯莱弗本人及天文台的另外两位资深天文学家小时候都有过类似的经历。

1929年1月，汤博乘坐了整整28小时的长途火车抵达罗威尔天文台，成为了天文台的一名观测助理。不久之后，在他的参与下，天文台的13英寸照相反射望远镜完成了装配及调试工作。

1929年4月，年轻的汤博正式走上了寻找海外行星的征途。

与发现天王星及海王星的时代相比，天文观测的手段，尤其是对暗淡天体的观测手段，已经有了很大的改善。早期的观测需要观测者对天体坐标进行手工记录，这对于观测暗淡天体来说是极为不利的。因为夜空中越是暗淡的天体，数量就越多。当所要观测的天体暗淡到一定程度时，需要排查的天体数量就会多到让手工记录成为不可承受之重。为了解决这一问题，天文学家们将照相技术引进到了天文观测之中。这样，手工记录的天体坐标就由相片所替代，而原先需要通过核对坐标来做的寻找新行星的工作，则可以通过对不同时间摄于同一天区的相片进行对比来实现。

罗威尔当年采用的就是这样的方法。这种方法免除了对天体坐标进行手工记录的麻烦，但并不意味着天文观测从此变得轻松了。事实上，在所要寻找的天体足够暗淡时，即便这样的方法也充满了困难。因为一张相片往往会包含几万甚至几十万个星体，对比排查的任务极其艰巨，几乎达到了肉眼不可能胜任的程度。而且需要对比的星体越多，就越容易因疏忽而丢失目标。为此，天文学家们又采用了一种新的仪器，叫做

闪视比较仪（blink comparator）。这种仪器的工作原理很简单，就是将需要对比的相片彼此叠合、快速切换。显然，位置或亮度发生过变化的天体将会在相片的切换过程中显示出跳跃或闪烁，从而变得很显眼。更有利的是，闪视比较仪还可以与光学放大系统结合在一起，进一步提高分辨率。有鉴于此，罗威尔天文台的天文学家们早在罗威尔还在世时，就曾多次建议罗威尔购买闪视比较仪，并在1911年罗威尔的生日派对上成功说服了罗威尔。

不过闪视比较仪的设想虽然高明，真正使用起来却不是一件容易的事情，因为对同一天区的两张相片只有在拍摄角度、曝光强度、胶卷冲洗等方面都保持高度的一致，才能获得良好的闪视比较效果。否则的话，连那些背景天体也会因为相片本身的人为差异而显示出变化。为了获得最佳的对比效果，汤博细心归纳了在不同天气条件下所需的曝光时间，并选出了一些明亮天体作为校正角度的参照点。他对每个天区都进行三次拍摄，以便从中选出两张最接近的相片进行对比。

26 寒夜暗影

汤博的搜索工作从接近罗威尔预言的巨蟹座开始。起初他只负责拍摄，闪视比较的工作则交由另一位天文学家进行。1929年4月11日，汤博的搜索工作刚刚进入第5天，就成功地拍摄到了新行星的倩影。19天后，他在对同一天区进行拍摄时再次将新行星摄入了相片。可惜的是，4月11日的照相胶片因天气寒冷而产生了裂缝，并且记录本身也因太接近地平线而受到了大气折射的干扰，进一步影响了质量。负责闪视比较的天文学家没能从数以万计的天体中发现这组记录，从而错过了一次可能的发现。这是继罗威尔时代的两次影像之后，新行星又一次躲过了罗威尔天文台的搜索。

几个月后，负责闪视比较的天文学家越来越忙于其他工作，很难抽出时间从事闪视比较，汤博便决定将这项工作接到自己手上。自那以后，他每个月用一半的时间从事观测，另一半的时间用来做闪视比较。由于相片上的天体实在太多⁽¹⁾，为避免数量压倒质量，汤博将每张相片都分割成很多小块，每块包含几百个天体。显然，这是一项高度重复，并且极其枯燥的工作。一般来说，检查几平方英寸的相片就会花去一整天的时间。当然，要说其中一点兴奋之处也没有，那倒也不是，时不时地汤博会看到一些变动的天体。不过，这时可不能高兴得太早，因为有很多鱼目混珠的天体会让人误以为找到了目标。事实上，汤博在每组相片中都会看到几十个那样的天体。可惜它们要么是变星⁽²⁾，要么是小行星、彗星或已知的行星，却没有一颗是新行星。这种“狼来了”的虚假天体见得多了，非但不能再带来兴奋，反而容易使人产生麻痹心理。但汤博始终保持着高度的敏锐和冷静，既不放过半点可疑之处，也从未作出过任何错误的宣告。

又过了几个月，一无所获的汤博决定不再以罗威尔的预言为参考，

毕竟他老人家的“预言”就像火星运河一样，口碑并不高，再紧盯下去有在一棵树上吊死的危险。作出了这一决定后，汤博将搜索范围扩大到了整个黄道面的附近，他的这一决定终结了罗威尔的预言对他搜索工作的帮助，因为这时的他已经走上了类似于巡天观测的道路。

1929年在繁忙的观测中悄然逝去，汤博在亚里桑那州寒冷高原的观测室里几乎沿黄道面搜索了一整圈。1930年1月，他的望远镜重新转回到了最初搜索过的天区。唯一不同的是，上一次是别人在帮他做闪视比较，而现在却是他本人在做。

1月21日，那个九一个多月前曾经落网，却在闪视比较时从网眼里溜走的暗淡天体再次出现在了汤博的相片上——当然，这时候虽然“天知地知”，汤博本人却还不知道。1月23日和29日，在高原寒夜的极佳观测条件下，汤博完成了对这一天区的第二和第三次拍摄。

2月15日，汤博开始检查后两次拍摄的相片。还是老办法，先分割，然后一片一片地进行闪视比较。2月18日下午4时，他在对比以双子座 δ 星为中心的一小片天区的相片时，发现了一个亮度只有15等的移动星体。

就像曾经无数次重复过的那样，汤博对这一天体进行了仔细的查验。45分钟之后，除新行星外的其他可能性逐一得到了排除，兴奋不已的汤博找到资深天文学家兰普朗德（Carl Lampland），告诉他自己终于找到了新行星。已在罗威尔天文台工作了28年的兰普朗德幽默地回答说他已经知道了，因为他注意到了一直忙碌着的闪视比较仪的声音突然停止，并变成了长时间的静寂。小伙子一定是发现了什么。

很快，天文台的几位资深天文学家与汤博一起冲进工作室，开始紧张地复查。经初步确认后，斯莱弗台长决定对这一天体先进行一段时间的跟踪观测，然后再对外公布。斯莱弗的这个决定既是出于谨慎，也暗藏着一些私心，因为他想利用这段时间积累观测数据，以便在接下来的新行星轨道计算中夺得先机。

在接下来的一个月的时间里，在天气许可的每一个夜晚，所有其他工作通通被抛到了爪哇国，罗威尔天文台把全部的观测力量都投入到了对新天体的观测之中。这时候，再没有什么任务能比曾被当成鸡肋的新行星观测更重要了。

1930年3月13日，罗威尔天文台正式对外宣布了发现新行星的消息。这一天是罗威尔诞辰75周年的日子。149年前，也正是在这一天，赫歇耳发现了天王星。

不久之后，罗威尔天文台的天文学家投票从来自全世界的候选名字中选出了新行星的名字：普卢托（Pluto），它是罗马神话中的地狱之神。说起来令人难以置信，首先提议这一名称的竟是英国牛津的一位年仅11岁的小女孩，她曾经学过经典神话故事并且很感兴趣，于是就提议用地狱之神命名这颗离太阳最远，从而最寒冷的新行星⁽³⁾。在中文中，这一行星被称为冥王星。



冥王星的发现让崛起中的美国科学界欣喜不已，在欧洲天文界垄断重大天文发现这么多年之后，幸运之神终于溜达到了美利坚，一些美国媒体兴奋地将新行星称为“美国行星”。但当时也许谁也不会想到，这个以地狱之神命名的新天体在天堂里待了76年之后，竟会从行星宝座上跌落下来，堕回“地狱”。

读者们也许还记得，汤博对冥王星的搜索，是从接近罗威尔预言的位置开始的，他曾经记录过冥王星的位置，只是未被认出。而当他正式发现冥王星的时候，他在黄道面附近完成了一整圈的搜索，又重新回到了起始时的天区。这表明冥王星的位置距离罗威尔的预言并不远。事实上，冥王星被发现时的位置与罗威尔1914年所预言的行星X在1930年初的位置只相差6°^[4]，这虽不像海王星的预言那么漂亮，却也不算太差。继海王星之后，天体力学似乎又一次铸造了辉煌。发现新行星的消息被宣布后的第二天，哈佛学院天文台台长沙普利（Harlow Shapley）在费城的一次小范围演讲被临时换到了一个大场地，因为他决定在演讲中加入有关新行星的消息。那一天，数以千计的听众挤满了演讲大厅。当久违了的罗威尔相片出现在投影仪上时，全场响起了雷鸣般的掌声。听众们用发自内心的掌声向这位已故的天文学家致敬。此情此景，因研究火星运河而遭冷遇的罗威尔若泉下有知，也当含笑了。

但是，冥王星的发现果真是继海王星之后天体力学的又一次伟大胜利吗？

注释

^[1] 汤博的每张相片平均约包含十六万个天体，对银河系中心方向拍摄的相片上则有多达一百万个天体。

^[2] 变星通常显示为亮度变化的天体，与移动天体明显不同。但有些变星在亮度变小后会因为比相片所能记录的最暗淡的天体还要暗，而从相片中消失，这样的变星在闪视比较时很像

是一颗移出（或移入）相片范围的移动天体。

[\[3\]](#) 普卢托（Pluto）这位地狱之神还被用于命名1934年发现的第94号元素钚（plutonium）。1945年8月9日，用这一元素制作的原子弹将日本城市长崎带入了地狱。

[\[4\]](#) 冥王星被发现时的位置距皮克林1928年修正后的行星O的位置也只差6°左右，不过皮克林的计算信誉太低，很少有人当真。

27 大小之谜

冥王星被发现之后，天文学家们很快就对它的轨道及大小进行了研究。在这两方面，冥王星都显现出很大的特异性。这其中轨道研究相对比较容易，短短几个月后就大体确定了主要的轨道参数，其中半长径约为39.5天文单位，椭率约为0.248，倾角约为 17.1° 。与其他八大行星相比，这是一个相当另类的轨道，它的椭率与倾角都是创纪录的。由于轨道椭率很大，冥王星有时甚至会比海王星离太阳更近，这种轨道交错现象在已知行星中是绝无仅有的。而由于轨道倾角很大，冥王星在多数时候都处在离黄道面较远的位置上，因而特别不易被发现。但幸运的是，汤博搜索冥王星的那段时间，恰好是冥王星离黄道面较近的时候。

冥王星的轨道参数虽然很快就被确定了，但确定它的大小——这个大小既是几何意义上的，也是质量意义上的——却向天文学家们提出了一个极大的挑战。因为人们很快就发现，无论用什么样的望远镜也无法让冥王星显示出行星应有的圆面。自望远镜问世以来，除了将小行星当成行星的那些年（参阅第7章）外，这种无法显示行星圆面的情形还从未发生过。当然，天文学家们对此倒也并非无心理准备，冥王星被发现时的亮度只有15等，比人们预期的暗淡得多⁽¹⁾，除非冥王星表面物质的反光率低得异乎寻常，否则这样的暗淡只能有一个解释：那就是冥王星比人们预期的小得多。

那么冥王星究竟有多小呢？天文学家们用了几十年的漫长时光才搞明白了答案。

由于无法观测到圆面，天文学家们惯用的通过几何手段确定行星直径的方法在冥王星这里遭到了滑铁卢，取而代之的是通过亮度间接推断直径这一不太可靠的方法。这一方法之所以不可靠，是因为行星的亮度与直径并不存在固定的关系。同样亮度的行星，若表面物质的反光率

高，它的直径就小；反之，若表面物质的反光率低，则直径就大。对于像冥王星那样遥远的新行星，当时的天文学家们并无任何办法确定其表面物质的反光率，因此虽然知道亮度，却无法准确估计它的直径。既然连直径都无法准确估计，对质量的估计自然就更困难了，因为后者还依赖于一个新的未知数：冥王星物质的平均密度。

虽然没有可靠的方法，天文学家们还是对冥王星的质量进行了粗略估计。1930-1931年间，天文学家们估计的冥王星质量约在0.1到1个地球质量之间。与现代数据相比，这是非常显著的高估。但即便是这些高估了的数据，也立刻就对罗威尔有关冥王星的“预言”造成了毁灭性的打击。读者们也许还记得，我们在第24章中曾经介绍过，罗威尔给出的行星X的质量约为地球质量的6.6倍。如果冥王星的实际质量只有0.1到1个地球质量，那它对天王星或海王星轨道的影响显然要远远小于罗威尔的计算，而罗威尔通过那种错误的影响对冥王星位置所作的反推则不可能是正确的。因此在冥王星被发现后不久，人们就已意识到，冥王星的发现并不是海王星神话的重演。冥王星在距罗威尔的预言只差 6° 的地方被发现，是纯粹的巧合⁽²⁾。

有读者也许会问：我们在第20章中曾经提到过，亚当斯与勒维耶对海王星质量及轨道的预测与海王星的实际参数也有不小的出入。为什么那些出入并不妨碍我们将海王星的发现视为巨大的天体力学成就呢？这首先是因为，亚当斯与勒维耶的海王星轨道计算是依据确凿存在的天王星出轨现象进行的，因此其观测依据是充分的；其次，20世纪七八十年代曾有人对亚当斯与勒维耶的计算细节进行了“复盘”，结果表明他们的计算细节也是完全有效的⁽³⁾。反观罗威尔有关冥王星的“预言”，虽然在计算方法上效仿了勒维耶，但它依据的所谓天王星与海王星的“出轨”数据是子虚乌有的，因而整个计算只是一场“空对空”的演练。另一方面，由于罗威尔的“预言”很快就被判定为无效，后人也就没兴趣去复核他的计算细节了，他在这方面犯错的可能性也是完全存在的。因此，

对冥王星的“预言”并不是海王星神话的重演，不仅在理论上不是，而且在实际上——如我们在上章中所说——也并未对冥王星的发现起到引导作用，冥王星的发现者汤博是在搜遍了黄道面之后才发现冥王星的。

虽然罗威尔有关冥王星的预言很快就被推翻了，但人们对冥王星大小的推算却仍在继续。直到冥王星被发现40年后的20世纪70年代初，人们对冥王星质量的估计仍大体维持在0.1到1个地球质量之间，这些估计与现代值相比都大得离谱。虽然推算冥王星的质量不是一件容易的事情，但在那么多年的时间里，那么多天文学家所作的那么多估算竟然一面倒地巨幅高估冥王星的质量，这其中不能说没有心理上的原因。这原因就是自木星开始，太阳系的外行星是清一色的巨行星，而冥王星又一经发现就被认定为是行星。虽然冥王星已绝无可能是巨行星，但天文学家们显然还没有足够的心理准备来接受有关它大小的真相。

我们在前面说过，同样亮度的行星，表面物质的反光率越低，相应的直径就越大。为了让冥王星维持一个体面的大小，天文学家们不惜将它“抹黑”为一个表面反光率极低、如同巨型煤球一样的天体。而事实上，在冥王星那样遥远而寒冷的行星上，很多气体都能凝结成冰，冥王星是一个具有较高表面反光率的“冰球”的可能性要比它是“煤球”的可能性大得多。这一显而易见的可能性被错误地蒙蔽了几十年，直到20世纪70年代中期，才终于被确立了起来。反光率的调整立即对冥王星的质量估算产生了巨大影响，它的质量估计值一举缩小了两个数量级，不仅比所有其他行星都小得多，甚至变得比月球还小。这也为它日后的命运沉沦埋下了种子。

不过，依靠对那样遥远的一个天体的表面反光率及物质密度的研究来推断其质量，无论如何只能算是下策。估计冥王星质量的最佳途径，显然是越过所有这些与冥王星物质有关的细节来直接估计其质量。这样的途径在1978年成为了现实。1978年6月22日，美国海军天文台（Naval Observatory）的天文学家们发现了冥王星的卫星卡戎（Charon，希腊神

话中摆渡亡灵的神）（图13）。在行星天文学上，一颗行星一旦被发现
有卫星，我们就可以通过观测卫星的运动来测定该行星的引力场，既而
推断其质量，这是测定天体质量最有效的手段之一。因此卡戎的发现为
直接估计冥王星的质量提供了极好的条件。（请读者们想一想，中学物
理课本中的哪一条定律有助于利用卡戎来确定冥王星的质量？） [\(4\)](#)



图13 从冥卫三看冥王星与卡戎（冥卫一）的艺术想像画

如今我们知道，冥王星的质量只有地球质量的0.21%（图14），它
绝不可能是罗威尔或其他任何人所预言的海外行星，它对天王星和海王
星的引力摄动甚至还不如作为内行星的地球对它们的引力摄动来得大。
1993年，美国加州喷气动力实验室的科学家斯坦迪什（Erland Myles
Standish, Jr）利用“旅行者号”飞船所获得的有关木星、土星、天王星和

海王星的最新质量数据重新计算了外行星的轨道摄动，并再次证实了的确不存在天王星和海王星的出轨问题，不存在需要用新行星来解释的偏差。冥王星的发现完全是一个多重错误导致的奇异果实：罗威尔对冥王星轨道的计算是依据错误数据所做的无效分析；汤博对冥王星的搜索则是源于罗威尔天文台对一个错误心愿的盲目继承。

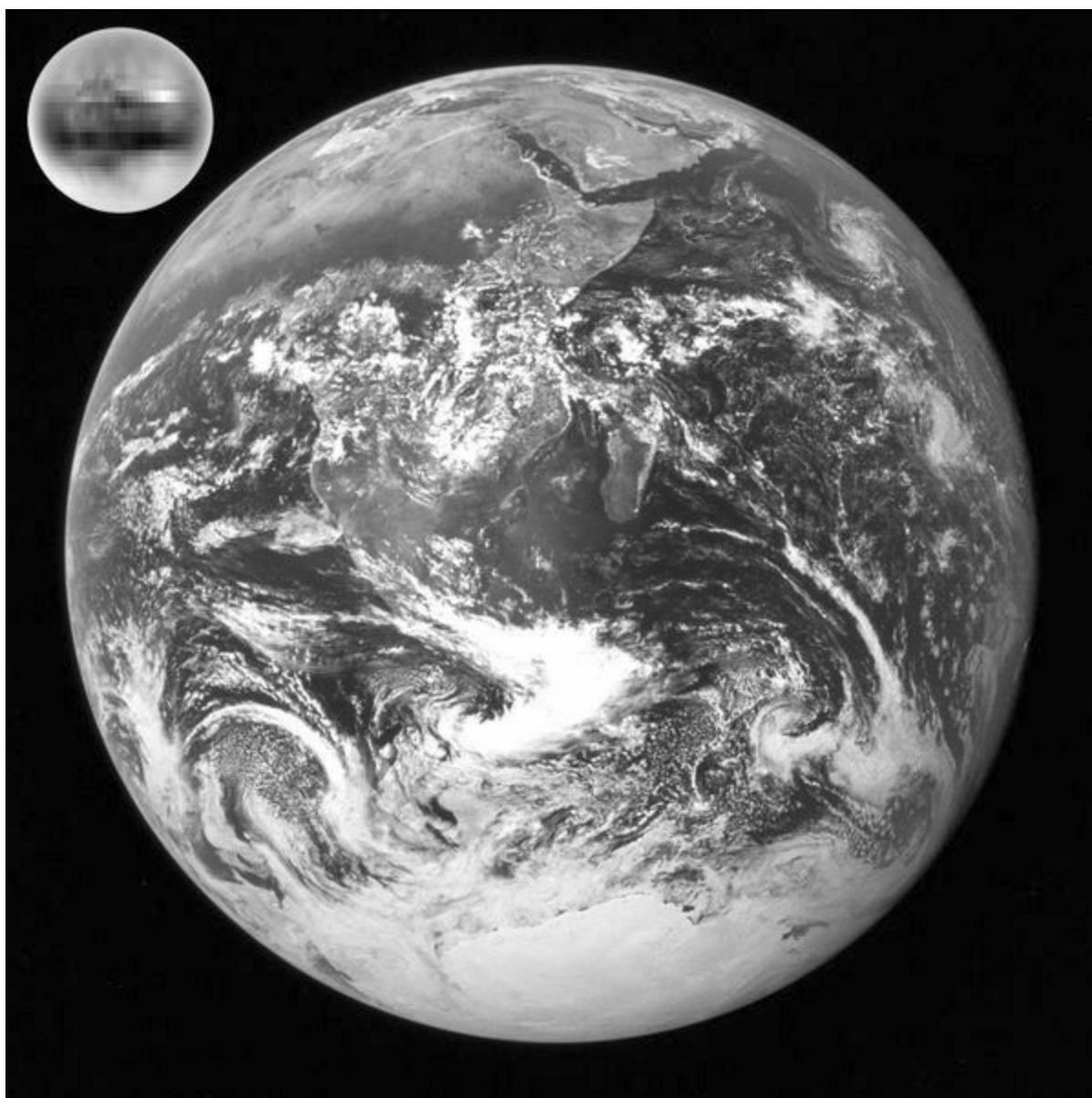


图14 冥王星（左上）与地球的大小对比

而所有这一切的错误之所以最终结出了一个如此美丽的果实，全靠

汤博在寒冷的亚里桑那高原上为期十个月的顽强搜索，这是整个冥王星故事中唯一的必然。

注释

[\[1\]](#) 比如罗威尔所预测的冥王星亮度为13等。

[\[2\]](#) 这一巧合的概率并不很小，因为罗威尔对行星X的位置预言其实有两处（彼此相差 180° ），在其中任何一处的左右 6° 范围之内发现新行星的概率约为 $1/15$ （请读者自行计算一下）。

[\[3\]](#) 1970年，一位名叫布鲁克斯（C. J. Brookes）的研究者对亚当斯的方法进行了分析，结论是它的确可以得到精度在几度之内的结果。1980年，另一位研究者巴格代迪（Baghdady）对勒维耶的方法进行了复盘，结果得到了误差仅为 $16'$ 的结果。这些验证表明亚当斯与勒维耶的计算方法都是有效的。

[\[4\]](#) 通过卡戎的运动直接测定的其实是冥王星与卡戎这一行星-卫星系统的总质量。对于其他行星来说，这几乎就等于行星的质量。但冥王星与卡戎却是一个引人注目的例外，因为卡戎的质量相当大（约为冥王星质量的11.65%）。因此用引力效应测定冥王星的质量时还牵涉到确定卡戎与冥王星的相对质量这一额外的复杂性。

28 深空隐秘

发现冥王星之后，汤博并未离开寻找太阳系疆界的孤独事业，他投入了另外13年的漫长时光，继续搜索更遥远的行星。他的搜索范围超过了整个夜空的2/3，他所涵盖的最低亮度达到了17等，他对比过的天体多达九千万个。在那13年里，他发现了6个星团、14颗小行星及一颗彗星，但却没能发现任何冥王星以外的新行星。

那么，冥王星轨道是否就是太阳系的疆界呢？既然观测一时还无法回答这个问题，天文学家们便展开了理论上的探讨。不过那探讨不再是像亚当斯与勒维耶那样的精密计算。由于冥王星的发现已属巧合，在那之后的天文学家们即使在做梦的时候，恐怕也很少会再幻想重演一次笔尖上预言新行星的奇迹了。但是，精密的预言虽不可能，粗略猜测一下太阳系的疆界在哪里却还是可以的。

那样的猜测几乎立刻就出现了。冥王星发现之初，美国加州大学的天文学家利奥纳德（Frederick C. Leonard）就猜测冥王星的发现有可能意味着一系列海外天体（trans-Neptunian object, TNO）将被陆续发现。应该说，在经历了天王星、海王星及冥王星的发现之后，单纯作出这样一个猜测已无需太高级的想象力了。不过，比单纯猜测更有价值的是，1943年爱尔兰天文学家埃奇沃斯（Kenneth Edgeworth）提出的稍具系统性的观点。

在介绍埃奇沃斯的观点之前，让我们稍稍介绍一下太阳系的起源学说。在科学上，几乎任何东西——人类、生命、地球乃至宇宙——的起源都是值得探究的课题，太阳系的起源也不例外。自18世纪康德

（Immanuel Kant）和拉普拉斯（Pierre-Simon Laplace）彼此独立地提出了著名的星云假说以来，天文学家们关于太阳系起源的主流观点是，太阳系是由一个星云演化而来的。这其中行星的形成，乃是来自于星云盘

上的物质彼此碰撞吸积的过程。

按照这种理论，行星形成过程的顺利与否与星云物质的密度有很大的关系。星云物质的密度越低，则引力相互作用越弱，星云盘上物质相互碰撞的几率越小，从而吸积过程就越缓慢，行星的形成也就越困难。当星云物质的密度低到一定程度时，行星的形成过程有可能缓慢到在太阳系迄今50亿年的整个演化过程中都无法完成，而只能造就一些“半成品”：小天体。埃奇沃斯认为，海王星以外的情形便是如此。那里的星云物质分布是如此稀疏，以至于行星的形成过程无法进行到底，而只能形成为数众多的小天体。由此他提出，人们将会在海王星之外不断地发现小天体，且那些小天体中的某一些会偶尔进入内太阳系，成为彗星。



美籍荷兰裔天文学家 柯伊伯（1905-1973）

无独有偶，1951年，美籍荷兰裔天文学家柯伊伯（Gerard Kuiper）也注意到了太阳系物质分布在海王星之外的急剧减少。与利奥纳德类似，他也认为那样的物质分布会形成一系列小天体而非大行星^{[\(1\)](#)}。但与利奥纳德以及后来的天文学家们不同的是，柯伊伯认为那些曾经存在过的小天体早已被冥王星的引力作用甩到了更遥远的区域，不会再存在于距太阳30~50天文单位的区域中了。换句话说，他认为在冥王星轨道的附近曾经有过大量的小天体，但目前已不复存在。在这点上，柯伊伯犯了一个可以原谅的错误，他以为冥王星的质量接近于地球质量（这在

当时被认为是有可能的），从而有足够的引力来做到这一点。而事实上，如我们在上章中介绍的，冥王星的质量只有地球质量的0.21%。

埃奇沃斯与柯伊伯的想法在接下来的十年间并未引起重视。但常言道：是金子总会发光的。一个合理的想法纵然一时沉寂，终究还是会复活的。1962年，在美国工作的加拿大天文学家卡梅伦（Alastair Cameron）提出了类似的看法。两年后，美国天文学家惠普尔（Fred Whipple）也加入了这一行列。惠普尔的研究比前面几位更加深入，除了猜测在海王星之外存在类似于小行星带的结构外，他还试图研究那些小天体对天王星和海王星轨道的摄动，但没能得到可靠的结果。1967年，惠普尔及其合作者又研究了七颗轨道延伸到天王星之外的彗星，试图寻找来自海外天体的引力干扰，结果也未发现任何可察觉的干扰。由此他们估计出那些小天体——如果存在的话——的总质量必定远小于地球质量。他们的这一估计在如今看来是颇有前瞻性的，但在当时却是一个有点令人沮丧的结果，因为它意味着观测那些小天体将会是一件非常困难的事情。

除了这些从太阳系起源角度所做的分析外，天文学家们从另一个完全不同的角度出发，也殊途同归地提出了海王星以外存在大量小天体的假说。这个不同的角度便是彗星的来源。彗星是太阳系中最令人瞩目的天体，当它们拖着美丽的尾巴（彗发）出现在天空中时，常常是万人争睹的天象。天文学家们注意到，太阳系中的彗星按轨道周期的长短大致可分为两类：一类是长周期彗星，它们的轨道周期在两百年以上，长的可达几千、几万、甚至几百万年。另一类则是短周期彗星，它们的轨道周期在两百年以下，短的只有几年。短周期彗星的存在给天文学家们带来了一个难题。因为这些彗星上能够形成彗发的挥发性物质会因频繁接近太阳而被迅速耗尽，而且它们的轨道也会因反复受到行星引力的干扰而变得极不稳定。计算表明，短周期彗星的存在时间应该很短，相对于太阳系的年龄来说简直就是弹指一瞬。但我们却在直到太阳系诞生50亿

年之后的今天仍能观测到不少命如蜉蝣般的短周期彗星，这是为什么呢？天文学家们认为，唯一的可能是太阳系中存在一个短周期彗星的补充基地。



这个短周期彗星的补充基地究竟在哪里呢？1980年，乌拉圭天文学家费尔南德斯（Julio Fernández）提出了一个后来被普遍接受的假说，即短周期彗星来自海王星之外的小天体带。他并且推测那些小天体的视星等约在17~18之间（比汤博曾经搜索过的天体更暗，但这个亮度后来被证实为仍是显著的高估）。在他颇具影响力的论文中，费尔南德斯援引了柯伊伯的文章，却忽略了埃奇沃斯的工作。费尔南德斯的这一粗心大意导致的后果是，人们多少有点乌龙地用柯伊伯的名字命名了那个小天体带。而事实上，如我们在上面提到的，在所有曾经猜测过那个小天体带的天文学家中，柯伊伯几乎是唯一一个认为它目前已不复存在——从而与费尔南德斯的假说及后来的观测结果截然相反——的人。费尔南德斯的假说提出之后，1988年，几位在美国加州大学及加拿大多伦多大学工作的天文学家通过计算机模拟手段，对这一假说进行了检验。他们的检验表明，由那样一个小天体带所产生的短周期彗星无论在数量还是轨道分布上都与实际观测有着不错的吻合。

因此，到了20世纪80年代末，来自不同角度的理论分析均表明，在海王星的轨道之外很可能存在一个小天体带，它是行星演化过程中的半成品，同时也是短周期彗星的大本营。但到那时为止，那个遥远的天区除了一颗孤零零的冥王星外，在观测意义上还是一片虚空。

距离给了外太阳系神秘的面纱，天文学家们却要揭开面纱来寻找隐秘。

注释

[\(1\)](#) 柯伊伯并未在自己的论文中提及埃奇沃斯的工作，这一点使得后来有历史学家对他是真的不知道埃奇沃斯的工作，还是暗中“借用”了对方的想法产生了疑问。

29 巅峰之战

在经历了追捕小行星的波折，发现海王星的纷争，搜寻火神星的未果，以及预言冥王星的虚无之后，在太阳系边缘搜索新天体的苦力活早已失去了往日的魅力。行星这个曾经神圣的概念渐渐变成了如美国物理学家费恩曼（Richard Feynman）在其名著《费恩曼物理学讲义》中所说的“那8个或10个遵循相同物理定律，由同样的尘埃云凝聚而成的球体”。在20世纪天文学发展的迅猛浪潮中，行星天文这个最古老的分支甚至一度整体性地沦落为了二流学科，以至于20世纪60年代，当美国国家航空航天局（NASA）为行星探测计划寻求咨询时，为天文学家们在这一分支上的知识贫乏而感到惊讶。后来，随着六七十年代美国与前苏联的一系列无人探测器计划的成功实施，行星天文学虽然重新成为了焦点领域，但与此同时，行星天文学家们的目光却也被吸引到了行星地貌、行星物理、行星化学等新兴方向上，对搜索新天体的兴趣依然低迷。

不过，当有关海外天体的猜测变得越来越言辞凿凿时，外太阳系的奥秘终于还是再次引起了一小部分天文学家的关注与喜爱。这其中麻省理工学院的一位天文学家决定化“爱心”为行动，展开对海外天体的观测搜索。这位天文学家名叫朱惠特（David Jewitt），来自英国。朱惠特七岁那年曾有幸目睹过一次流星雨，年幼的他被天象的美丽与神奇所吸引。20世纪70年代后期，美国国家航空航天局发布的美轮美奂的行星及卫星图像再次打动了当时正在伦敦念本科的朱惠特。他决定选择行星天文学作为自己的专业，并前往美国念研究生。1983年，朱惠特在美国加州理工大学获得了博士学位，随后成为了麻省理工学院的助理教授。在那里，他遇到了重要的学术合作伙伴卢简（Jane Luu）[\[1\]](#)。卢简是一位出生于越南的女孩，1975年随父母逃难来到美国。与朱惠特一样，卢

简也是被美国国家航空航天局的行星与卫星图像所吸引，而选择了行星天文学作为自己的专业。朱惠特在麻省理工学院的时候，卢简正在那里念研究生。

1987年的某一天，当朱惠特和卢简在系里相遇时，朱惠特提议卢简参与自己即将开始的搜索海外天体的工作。这是自冥王星被发现之后将近半个世纪的时间里极少有人问津的冷门观测。卢简问朱惠特：“为什么要做这样的观测？”朱惠特的回答是：“如果我们不做，就没人做了。”听起来颇有几分“我不入地狱，谁入地狱”的悲壮。卢简被这个简短的回答所打动，一场历时五年的漫长搜索由此揭开了序幕。

朱惠特与卢简最初的观测地点是位于亚里桑那州的美国基特峰国家天文台（Kitt Peak National Observatory）及南美洲的塞罗托洛洛天文台⁽²⁾，他们最初采用的观测方法类似于汤博当年所用的方法，即通过对间隔一段时间拍摄的同一天区的相片进行闪视比较，来寻找缓慢运动的天体。当然，半个世纪之后的朱惠特与卢简所拥有的设备已非汤博当年可比，唯一不变的是任务本身的繁重、枯燥，以及用眼过度产生的疲惫。经过了一段时间的搜索，朱惠特与卢简一无所获，他们辛苦寻获的运动天体无一例外地被证实为是已知天体、胶片缺陷、灰尘或宇宙线造成的影像。

幸运的是，就在这时，一项让整个光学观测领域脱胎换骨的新兴技术——电荷耦合器件（Charge Coupled Device, CCD）——进入了天文界。CCD是1969年由美国贝尔实验室（Bell Labs）的两位科学家发明的、一种可以取代传统胶片的感光器件。CCD的最大优点是具有极高的敏感度，能对70%甚至更多的入射光作出反应，而普通照相胶片的这一比例还不到10%。真是不比不知道，一比吓一跳。要知道朱惠特与卢简所寻找的是离太阳几十亿千米之外的小天体，它们自身并不发光，全靠其表面反射的太阳光才能被我们所发现。在那样遥远的距离上，太阳的光芒只有约一亿亿分之一能够照射到那些小天体上。那部分光线有的被

吸收，有的被反射，那些反射光必须再次穿越广袤的行星际空间，其中只有约一万亿分之一能够来到地球。而在那“亿里迢迢”来到地球的反射光中，恰好能进入望远镜的又只有其中的约一百万亿分之一。这是何等宝贵的“星星之火”？可这宝贝却还要被该死的照相胶片忽略掉90%以上，这真是“生可忍，熟不可忍”（韦小宝语）。

因此CCD的使用对于观测天文学来说堪称是一场革命。不过CCD虽然在感光性能上遥遥领先于普通胶片，在一开始却也有一个很大的缺陷，那就是像素太少。朱惠特与卢简最初使用的CCD的有效像素仅为 242×276 ，相当于如今一台普通数码相机像素数量的1%。由此带来的后果是，每张CCD相片涵盖的天区面积只有他们以前所用的普通光学相片的千分之一。换句话说，原先分析一组相片就能覆盖的天区，如今却要分析一千组相片。但幸运的是，CCD所采用的独特的感光方式为计算机对比相片开启了方便之门，从而大大减轻了对肉眼的依赖。而更重要的是，对于特别暗淡的天体，普通胶片有可能因为敏感度不够而无法记录，这时CCD的优势更是无与伦比。因此，当CCD进入天文观测领域后，朱惠特与卢简便决定用它取代照相胶片。

这时候，朱惠特与卢简的观测地点也发生了变化。1988年，朱惠特接受了夏威夷大学天文研究所的一个职位。不久，卢简也来到了夏威夷，两人利用夏威夷大学所属的茂纳基雅天文台（Mauna Kea Observatory）（图15）的一台口径2.24米的望远镜继续他们的海外天体搜寻工作。茂纳基雅是夏威夷语，含义是“白山”，那里常年积雪，而茂纳基雅天文台的所在之处正是白山之巅，海拔高达4200米（比汤博所在的罗威尔天文台高了一倍）。那里的空气稀薄而干燥，氧气的含量只有海平面的60%，常人在那里很容易出现高原反应，大脑的思考及反应能力也会明显下降。为了减轻高原反应的危害，天文学家们像登山者一样，在海拔较低（3000米）的地方建立了营地。要去天文台的天文学家通常提前一晚就来到营地过夜，以便让身体提前适应高原的环境，然后

在第二天晚饭之后驾驶越野车前往天文台。在那里，朱惠特与卢简夜复一夜地进行着观测。当他们感到疲惫的时候，有时朱惠特会放上一段重金属音乐，有时则卢简会放上一段经典音乐，控制室里响彻着时而激昂、时而舒缓的乐曲。



图15 茂纳基雅峰上的观测台

这样的日子一晃就是四年，其间卢简完成了自己的学业，并获得了哈佛大学的博士后职位，但她仍时常回到茂纳基雅天文台，与朱惠特一起，在那白山之巅的稀薄空气里继续着对海外天体的执著搜索。尽管一次次的努力换来的只是一次次的失望，但他们锲而不舍地坚守着这份孤独的事业。幸运的是，在那四年中，CCD的技术有了长足的发展，分辨率由最初的 242×276 提高到了 2048×2048 ，从而大大提升了搜索效率。在毅力、耐力和技术这三驾马车的共同牵引下，朱惠特与卢简这场巅峰之

战的胜利时刻终于来临。



1992年8月30日，在对比两张CCD相片时，一个缓慢移动的小天体引起了朱惠特的注意。一般来说，距离太阳越远的天体运动得越慢，从那个天体的移动速度来看，它与太阳的距离似乎有60天文单位。换句话说，这似乎是一个海外天体。当然，仅凭两张相片的对比是不足以作出结论的，于是他们对该天区进行了反复的拍摄与对比，结果证实这一天体的确是在缓慢地运动着，而且其运动速度所显示的距离的确是在海王星轨道之外，因此的确是一个海外天体。

朱惠特与卢简终于成功了。四年了，他们在这仿佛伸手便可摘到星星的巅峰之上苦苦寻找，运气却仿佛远在星辰之外。没想到成功竟然就在今夜，这一刻真让人猝不及防！朱惠特与卢简兴奋得像两个大孩子一样在观测室里又蹦又跳。他们将这一消息通告了国际天文联合会

（International Astronomical Union）所属的小天体中心（Minor Planet Center）^[3]。9月14日，小天体中心的天文学家马斯登（Brian Marsden）正式公布了这一发现，并确定了该天体的临时编号：1992QB₁^[4]。据测定，1992QB₁的轨道半长径约为44天文单位（比朱惠特最初估计的要小，但的确是在海王星轨道之外），直径约为160千米。

注释

^[1] 按照用姓氏称呼外国人名的惯例，Jane Luu应该被称为卢，考虑到一个字的中文名用起来比较别扭，本书将Jane Luu按全名译为卢简。

^[2] 我们曾在第20章中提到过这个天文台，海王星档案就是在那里失而复得的。塞罗托洛天文台虽远在智利，却是美国国家光学天文台（National Optical Astronomy Observatory）的一部分。

^[3] Minor Planet Center若直译，应为“小行星中心”，但考虑到中文的“小行星”一词往往特指由英文asteroid所表示的小行星带中的小天体，因此本书将之译为“小天体中心”。

(4) 自1925年以来，天文界采用了以发现年份外加两个英文字母作为小天体临时编号的做法。其中第一个字母（I与Z不出现）表示发现小天体的半月，从一月上半月的A到十二月下半月的Y。第二个字母（I与Z同样不出现）则按照小天体在该半月中的发现顺序排列。如果该半月中发现的天体数目超过24个，则以下标表示字母被重复使用的次数。请读者按照这一命名规则推算一下1992QB₁是哪一个半月发现的？以及它是该半月中被发现的第几个小天体？

30 玄冰世界

1992QB₁的发现是人类在寻找太阳系疆界的征途上取得的又一个重要进展。不过在一开始，有些天文学家对1992QB₁是否真的是海外天体还心存疑虑。比如小天体中心的马斯登，他虽然亲自宣布了1992QB₁被发现的消息，但其本人却是怀疑者中的一员。他认为1992QB₁有可能只是一个轨道椭率很大的天体，这样的天体虽然远日点距离很大，但绝大多数时间其实都处在海王星轨道以内，从而算不上是货真价实的海外天体。马斯登甚至为自己的猜测与朱惠特打了500美元的赌。

这个赌局很快就有了结果。1993年3月28日，朱惠特与卢简发现了第二个海外天体，临时编号为1993FW。1993FW的轨道及大小都与1992QB₁相似，它的发现极大地动摇了马斯登的怀疑，因为天文学家们在对这两个天体的轨道计算中犯下同样错误，一错再错地把轨道椭率很大的天体误当成海外天体的可能性是很小的。此后不久，更多的海外天体被陆续发现，从而越来越清楚地表明它们正是理论家们几十年前所猜测的那个海外小天体带的成员。1994年，当海外天体的数量增加到六个（其中四个是朱惠特与卢简发现的）时，马斯登终于“投降”，乖乖交出了500美元。

与当年发现小行星带的情形相类似，随着观测技术的持续改进，以及受第一轮发现的吸引而对海外天体感兴趣的观测者的增多，海外天体的发现不断提速，在热闹的年份里一年就能发现一两百个（当然，它们的发现也因此而很难再登上新闻标题了）。不过，由于距离遥远，加上体形苗条，海外天体大都极其暗淡，视星等通常在20以上，不到冥王星被发现时的亮度的1%；加上观测海外天体在各大天文台的任务排行榜上的地位较低，因此被发现的海外天体因未能及时跟踪而重新丢失的比例也大得惊人，有时竟达40%。寻找海外天体的努力，仿佛是往小学数

学题里那个开着排水口的水池里灌水，一边找，一边丢。不过在一群像朱惠特与卢简那样执著的天文学家的努力下，得到确认的海外天体（图16）的数量还是稳步增长着。截至2008年3月，被小天体中心记录的海外天体数量已经超过了1300，它们的表面大都覆盖着由甲烷、氨、水等物质组成的万古寒冰。

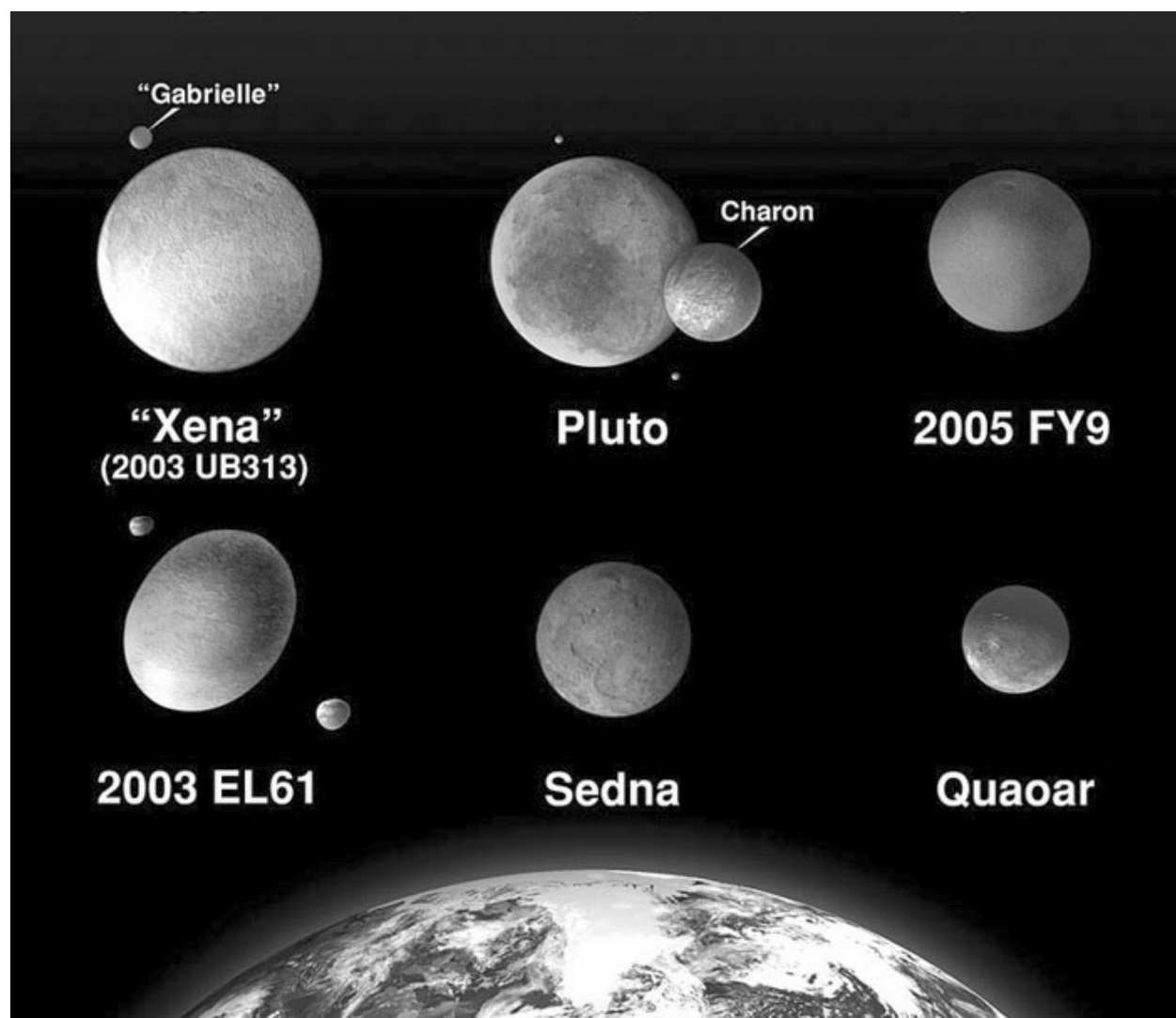


图16 若干直径较大的海外天体（下方为地球）

随着数量的增加，天文学家们对海外天体按其轨道特征进行了粗略的分类，其中距太阳30~55天文单位的海外天体被称为柯伊伯带天体（Kuiper belt object），它们构成了所谓的柯伊伯带。我们在第28章中

曾经提到，“柯伊伯带”这一名称其实有点乌龙，因为在曾经猜测过这一小天体带的天文学家中，柯伊伯的观点偏偏是认为它们如今早已不复存在，从而与观测结果完全不符。不过，柯伊伯是一位对现代行星天文学有过重要影响，甚至被一些人视为现代行星天文之父的天文学家，用他的名字命名一个天体带也不算过分。据估计，柯伊伯带中直径在100千米以上的天体可能有几万个之多，目前已被发现的还只是冰山之一角。

另一方面，柯伊伯带天体相对于全部海外天体来说也同样只是冰山之一角。在发现柯伊伯带的过程中，人们也发现了一些离太阳更远的天体，那些天体被称为离散盘天体（scattered disc object），它们的轨道椭圆率通常很大，轨道倾角的范围也比柯伊伯带天体宽得多，它们的远日点比柯伊伯带天体离太阳远得多，但近日点却往往延伸到柯伊伯带，个别的甚至会向内穿越海王星轨道。一般认为，离散盘天体最初也形成于柯伊伯带之中，后来是因为受到外行星的引力干扰而被甩离了原先的轨道。有鉴于此，天文学家们有时将离散盘天体称为离散柯伊伯带天体（scattered Kuiper belt object）[\(1\)](#)。

人们早期发现的海外天体的直径大都在一两百千米左右，但渐渐地，一些更大的天体也被陆续发现了。（请读者想一想，哪些因素有可能导致那些更大的海外天体反而较迟才被发现？）下表[\(2\)](#)列出了其中较有代表性的几个（其中“正式编号”是小天体中心在轨道被确定后指定给小天体的编号）：

正式编号	临时编号	名 称	直径/千米
19308	1996TO ₆₆		~ 900
20000	2000WR ₁₀₆	Varuna	780 ~ 1016
55565	2002AW ₁₉₇		890 ~ 977

续表

正式编号	临时编号	名 称	直径/千米
50000	2002LM ₆₀	Quaoar	1200 ~ 1290
84522	2002TC ₃₀₂		710 ~ 1200
136108	2003EL ₆₁	Haumea	1200 ~ 2000
90482	2004DW	Orcus	909 ~ 1500

这些天体的大小都接近或超过了最大的小行星——谷神星（谷神星的直径约为960千米）。看来这遥远的玄冰世界里还真是别有洞天。不过，这些天体与行星世界的小弟弟，直径约2300千米的冥王星相比终究还是偏小了一点。

但就连这一点也在2005年的新年伊始遭遇了挑战。

2005年1月5日，美国加州理工大学的行星天文学家布朗（Michael Brown）在检查一年多前（2003年10月21日）他与北双子天文台（Gemini North Observatory）的天文学家特鲁吉罗（Chad Trujillo）及耶鲁大学的天文学家拉比诺维茨（David Rabinowitz）拍摄的相片时，发现了一个新的海外天体。按照相片拍摄的时间，这一天体的编号被确定为2003UB₃₁₃。2003UB₃₁₃是一个轨道椭率很大的天体，它被发现时正处于距太阳约97.5天文单位的远日点。在那样遥远的距离上仍能被观测到，可见其块头一定小不了。据布朗估计，2003UB₃₁₃的直径起码比冥王星大25%[\(3\)](#)。这一估计在行星天文学界引起了很大的震动。因为自冥王星被发现以来，这还是人们首次在太阳系中发现比冥王星更大，同时又不是卫星的天体。毫无疑问，像2003UB₃₁₃那样的庞然大物应该有一个专门的名称，它曾被暂时命名为齐娜（Xena），后来被正式定名为埃里斯（Eris）。这是希腊神话中的争吵女神，著名的特洛伊之战（Trojan war）就是在她的煽风点火之下引发的。在中文中，这一天体被称为阋神星。

这位不太淑女的女神很好地预示了她即将带给天文学家们的东西：争吵，有关行星定义的争吵。

注释

[\[1\]](#) 离散柯伊伯带天体还包括所谓的半人马小行星（centaurs），那也是一些轨道偏心率很大的小天体，只不过与离散盘天体的向外离散恰好相反，它们是向内离散的，其轨道通常分布于木星轨道与海王星轨道之间。

[\[2\]](#) 表格中的数据是早期的估计值，大都有些偏高。天文学家们一直在对海外天体的大小进行观测和修正，比如20000Varuna的直径后来（2007年）通过斯皮策太空望远镜（Spitzer Space Telescope）的观测而被修正为500千米左右。

[\[3\]](#) 这一估计有些偏高，目前人们对2003UB₃₁₃直径的估计为（2400±100）千米，只比冥王星略大，不过它的质量要比冥王星大28%左右，这一点由于它与冥王星分别存在卫星而得到了比它们的直径对比更为可靠的确立。

31 冥王退位

阋神星的发现向天文学家们提出了一个问题，那就是：它究竟是不是行星？这原本不应该成为问题的，因为阋神星既然比冥王星还大，当然应该算是行星。但问题是，在阋神星之前，人们已经发现了大量的海外天体，并且已经接受了海外天体是行星演化过程中的半成品的想法。在这种背景下要接受阋神星为行星是有难度的。更何况，海外天体中还包含了其他一些大小可观的成员。除上章列出的夸欧尔（Quaoar，美国原住民神话中的创世之神，中文名称为创神星），好瑟妹阿

（Haumea，美国夏威夷神话中掌管生育的女神，中文名称为妊神星）及厄耳枯斯（Orcus，罗马神话中的死亡之神，中文名称为死神星）等外，还有与阋神星同一天被宣告发现的马克马克（Makemake，复活节岛上的造物之神，正式编号为136472，发现时的临时编号为2005FY₉，中文名称为鸟神星），它的直径也有1300~1900千米。这些天体虽比冥王星小，但相差并不多，如果阋神星和冥王星可以算作是行星，那它们是否也应该算是行星呢？

当人们开始提出这样的问题时，一个更基本的问题也随之浮出了水面：究竟什么是行星？



就像其他很多习以为常的概念一样，人类知道行星的存在虽有漫长的历史，却从未给它下过明确的定义。在历史上，人类对行星的认定极少发生争议，而且即便发生争议，也要么很快就被解决（比如有关小行星地位的争议），要么所争之处并非行星的定义（比如对地球地位的争议），从而并未触及行星定义的必要性。

可现在的情况完全不同了。要知道冥王星行星资格的由来有着很大的偶然性：它一开始就被错误地当成了罗威尔的行星X，可以说是将行星宝座当成婴儿床，直接就诞生在了那里。尔后又在很长的时间内被误认为可能有地球那么大。后来虽一再“瘦身”，但生米早已煮成熟饭，再说“瘦死的骆驼比马大”，冥王星虽小，比小行星终究还是大得多，因此其身份虽遭到过怀疑，却像一位有经验的潜伏人员那样有惊无险地挺了过来⁽¹⁾。但随着海外天体的陆续登场，冥王星除在个头上遭到挑战外，它隐匿多年的一桩“劣迹”也得到了曝光。我们知道，当年小行星们之所以被剥夺行星资格，除个头太小外，还因为它们犯有一项“重罪”，那就是“非法聚众”。现在冥王星显然也犯下了同样的“罪行”。在这种情况下，摆在天文学家们面前的是一个两难局面：要么像当年处理小行星一样，剥夺冥王星的行星资格；要么一视同仁地将所有较大的柯伊伯带天体全都吸收为行星，甚至恢复某些小行星的名誉。无论哪一种选择，都将改变已沿袭了大半个世纪的太阳系九大行星的基本格局。

另外需要提到的是，除了来自太阳系内部的这些麻烦外，行星这个被太阳系垄断了几千年的专利，自20世纪90年代开始遭遇了“盗版”。天文学家们在其他恒星（包括白矮星、脉冲星等恒星“遗体”）周围也陆续发现了行星，而且其数目迅速增加，目前已远远超过了太阳系的行星数目。所有这些都促使天文学家们摆脱单纯的历史沿革，对行星的定义进行系统思考。在这过程中，冥王星的命运是让很多人——尤其是公众——最为关注的焦点。

1999年，随着有关冥王星地位变更的传闻越来越多，负责天体命名及分类的国际天文联合会发表了一份声明，公开否认其正在考虑这一问题。但就在这份明修栈道式的声明发表的同一年，该联合会却暗度陈仓般地成立了一个旨在研究太阳系以外行星（Extrasolar Planet）的工作组。2001年2月，该工作组拟出了一份名义上只针对太阳系以外行星的定义草案，其中给出了行星定义的一个重要组成部分，那就是行星必须足够小，以保证其内部不会发生核聚变反应⁽²⁾。这一条的主要目的是将行星与所谓的褐矮星（brown dwarf）区分开来。按照我们目前对天体内部结构的了解，这一条给出的行星质量上界约为木星质量的13倍。

除上界外，完整的行星定义显然还应包含一个合理的下界，否则环绕恒星运动的任何小天体，甚至每一粒尘埃都将变成行星，那是不堪设想的事情。不过由于早期发现的太阳系以外的行星大都是巨行星，因此上述草案并未对质量下界给予认真关注，只是建议参照太阳系行星的情况。可这“参照”二字说来容易，做起来却绝不轻松，因为太阳系行星的情况一向只是约定成俗，而从未有过明确定义，若当真遇到什么棘手的情形，还真不知该如何参照。有鉴于此，2002年，美国西南研究所

（Southwest Research Institute）的天文学家斯特恩（Alan Stern）与莱维森（Harold Levison）提出了一组新的行星定义，这一定义采用了与上述草案相同的质量上界（措词略有差异），但补充了质量下界。它规定：行星必须足够大，以至于其形状主要由引力而非物质中的其他应力所决定。在太阳系中，我们可以看到很多形状不规则的小天体，但几乎所有直径在400千米以上的天体，其形状都非常接近由引力所主导的天然形状：球形⁽³⁾。因此由这一条给出的行星直径下界约为400千米，具体的数字则与天体的物质组成有关。

由上述方式定义的行星质量上界及下界具有非常清晰而自然的物理意义。有了这两条，再加上行星必须环绕恒星运动，以及行星不能同时是卫星这两个显而易见的运动学要求，行星定义就基本完整了。2006年

8月16日，国际天文联合会正式提出了一份行星定义草案。该草案所采用的大致就是上述几条，不过在涉及质量上界时，只对行星与普通恒星作了区分，而未涉及与褐矮星的区分（这相当于将质量上界由木星质量的13倍提高到75倍左右）。这份定义草案单从物理角度讲是比较令人满意的，但用到太阳系中却立刻会产生一个很现实的麻烦，即导致行星数量的急剧增加。事实上，由于该定义所要求的行星直径的下界只有400千米左右，一旦被采用，则不仅谷神星可以“官复原职”，阋神星能够“荣登宝座”，许许多多甚至连名字都还没有的家伙也将成为行星。据估计，这一定义有可能会使太阳系的行星数目增加到几百，甚至几千。这样的数目虽然不存在任何原则性的问题，却有点超乎人们的心理承受力，因为自冥王星被发现以来，几乎每一位小学生都能说出太阳系九大行星的名称。但假如九大行星变成几百、甚至几千大行星，那么别说小学生，恐怕连大学教授也得张口结舌。

因此，上述草案一经提出立刻遭到了激烈的反对。经过几天的争论，国际天文联合会在草案中新增了一项要求：行星必须扫清自己轨道附近的区域^{[\(4\)](#)}。2006年8月24日，这一新定义经表决以超过90%的大比率通过，从而正式生效。按照新增的那项要求，谷神星“官复原职”的希望付诸东流，阋神星“荣登宝座”的美梦也化为了泡影，而最惨的则是已经在行星宝座上端坐了76年的冥王星，它在一夜之间就被扫地出门，变成了所谓的“矮行星”——这是为像它这样满足其他各项要求，却没能完成轨道“大扫除”任务的天体所设的安慰奖。与冥王星一同获得首批矮行星光荣称号的还有谷神星和阋神星。2008年3月和9月，鸟神星和妊神星也先后加入了矮行星的行列。今后，矮行星的数目显然还会增加，但太阳系行星的数目却暂时降为了八个。也许是意识到新定义的修改过程太过仓促，国际天文联合会将新定义的适用范围限定在了太阳系以内，而将普遍的行星定义留给了未来。

行星新定义的仓促出炉，尤其是冥王星像“严打”期间遭到惩处的人

犯一样在几天之内就被草率“矮化”，引起了很多人的反对，反对者从天文学家到天文爱好者，从普通民众到占星术士应有尽有。以前太阳系有九大行星时，人们曾用九大行星的英文开首字母编写过一些便于记忆的英文短句，比如：My Very Educated Mother Just Served Us Nine Pizza（我那受过良好教育的妈妈刚给我们做了九个比萨饼），冥王星（“P”luto）被剥夺行星资格后，有人戏谑般地用剩下的八个开首字母也编了一个英文短句：Most Vexing Experience, Mother Just Served Us Nothing（最气恼的经历，妈妈没给我们做任何东西）。

当然，也有比较认真的反对者，比如有人对表决的代表性提出了质疑。他们指出，参与行星定义表决的天文学家只有424人（其中投反对票者为42人），不到与会人数的16%，与国际天文联合会的会员总数相比，更是连5%都不到，不能充分地代表国际天文联合会。不过这种质疑初看起来颇有说服力，其实却不然。因为国际天文联合会的会员并非人人都对行星定义感兴趣，因此投票率的高低未必能衡量投票质量的好坏。另一方面，424人从统计学角度讲已经不算是太小的样本，统计误差只有百分之几，超过90%的大比率通过绝非统计误差所能干扰。除非有迹象表明未投票的天文学家看待行星定义的态度与已投票者存在系统性的差异，否则更多的人投票只会使赞成及反对的票数大致按比例增加，却几乎不可能改变投票结果。

当然，最重要的是，行星定义无论如何改变，所影响的只是我们对天体的称呼与分类，而不是天体本身。冥王星是行星也好，是矮行星也罢，它就是那个在六十亿千米之外围绕太阳运动，直径约2300千米，“遵循相同物理定律，由同样的尘埃云凝聚而成”的实心球。它是否被新定义所“矮化”，无论对于它自己还是对于天文学研究都没什么实质意义。不过，如果读者对名分问题感兴趣的话，朱惠特——他曾被认为是最早发现柯伊伯带天体的天文学家，但现在只能排第二了（请读者想一想，第一是谁？）——倒是早在冥王星被“矮化”之前就表达过一个别

致的看法，他认为冥王星如果变成一个柯伊伯带天体，非但不是被“矮化”，反而是受到“升迁”，因为它的地位将从此“由外太阳系的一个令人难以理解的畸形反常，变成海外天体这一丰富而有趣的家族的首领”。正所谓：宁为鸡头，不做凤尾，看来我们应该祝贺冥王星^{[\(5\)](#)}。

注释

^{[\(1\)](#)} 对冥王星身份的最早怀疑可以追溯到汤博发现冥王星的同一年，即1930年，起因是罗威尔天文台公布的冥王星轨道与罗威尔对行星X的预言不符。

^{[\(2\)](#)} 确切地讲，该定义要求行星的质量小于在其中心产生氦核聚变所需的质量。由于氦核聚变是恒星内部最容易产生的核聚变，因此满足这一条也就自动保证了行星内部不会产生其他核聚变。

^{[\(3\)](#)} 确切地讲是椭球形，因为多数天体存在自转。

^{[\(4\)](#)} 这一条与其他几条相比，其缺陷是显而易见的，因为它并未对“轨道附近的区域”及“扫清”这两个概念进行界定。严格追究的话，海王星也不能算是扫清了轨道附近的区域，因为很多海外天体的轨道周期性地穿越海王星轨道。甚至最有行星资格的木星，它的“大扫除”也是有死角的，因为在它的轨道区域中存在数量多达十万以上的所谓“特洛伊小行星”（Trojan asteroid）。从国际天文联合会对新定义的讨论过程及此前出现的几篇相关论文来看，“扫清”一词的含义应该是指行星在其轨道附近的区域中处于支配性（dominant）地位。

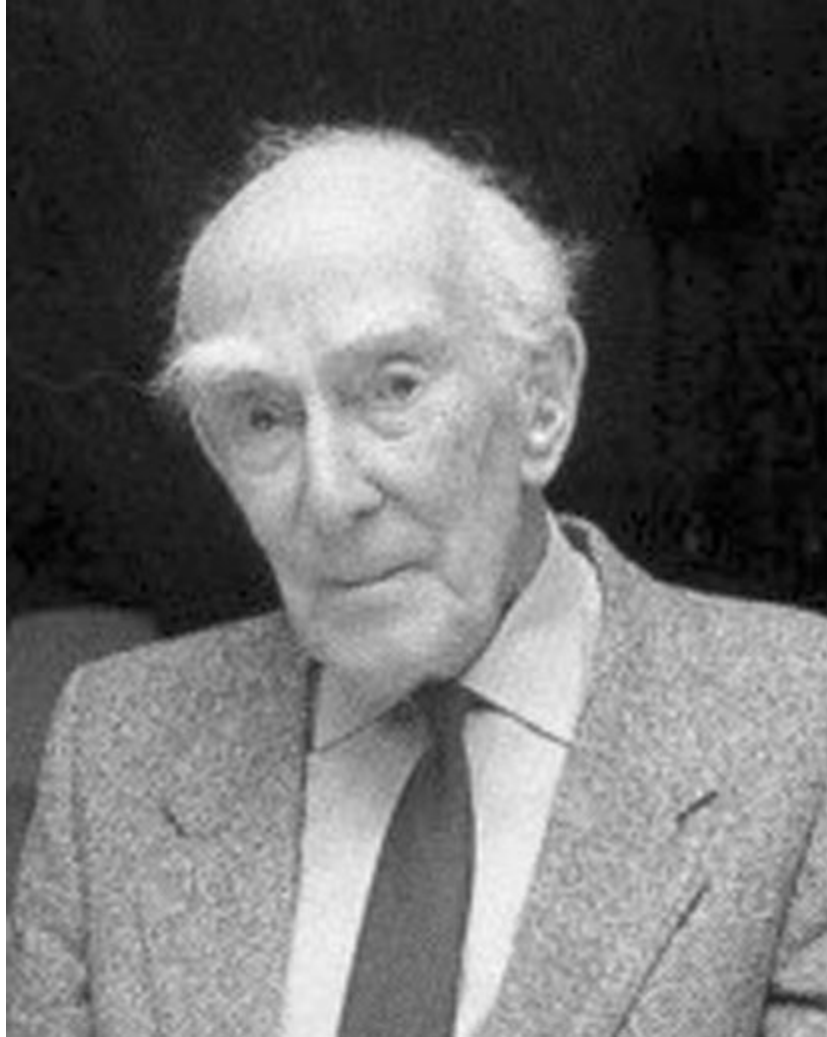
^{[\(5\)](#)} 冥王星的“鸡头”地位在2008年6月11日得到了进一步的加强：这一天，国际天文联合会将海王星以外（即轨道半长径大于海王星轨道半长径）的矮行星统称为[[WTHZ](#)]Plutoid[[WTBZ](#)]。该类别目前尚无标准中文译名，几种可能的选择为：类冥天体、类冥矮行星、冥王星类天体。其中个别译名曾被当作plutino——即与海王星轨道存在3：2共振的海外天体（包括卫星）——的非正式中译名。不过plutoid这一新类别出现后，为对两者进行区别，我认为plutino宜另找一个可以体现英文词根-ino（微小）的新词作为译名，比如类冥小天体、微冥天体等。

32 疆界何方

现在让我们盘点一下人类在寻找太阳系的疆界时走过的漫漫长路。从远古时期就已知道的金、木、水、火、土五大行星，以及脚下的地球，到近代的天王星、海王星，再到现代的柯伊伯带及离散盘。人类认识的太阳系疆界在过去两百多年的时间里在线度上扩大了十倍左右。

那么，离散盘是否就是太阳系的疆界呢？答案是否定的。

读者们也许还记得，我们在第28章中曾经提到，太阳系里的彗星按轨道周期的长短可以分为两类，其中短周期彗星大都来自柯伊伯带。那么，长周期彗星又来自何方呢？



荷兰天文学家奥尔特 （1900-1992）

1950年，荷兰天文学家奥尔特（Jan Oort）对长周期彗星进行了研究。他发现，很多长周期彗星的远日点位于距太阳50000~150000天文单位（约合0.8~2.4光年）的区域内，由此他提出了一个假设，即在那里存在一个长周期彗星的大本营。这一假设与将柯伊伯带视为短周期彗星补充基地的假设有着异曲同工之妙（但时间上更早）。那个遥远的长周期彗星大本营后来被人们用奥尔特的名字命名为奥尔特云（Oort Cloud）[\(1\)](#)（图17）。由于长周期彗星几乎来自各个方向，因此奥尔特云被认为大体上是球对称的。后来的研究者进一步将奥尔特云分为两部分：距太阳20000天文单位以内的部分被称为内奥尔特云，它呈圆环

形分布；距太阳20000天文单位以外的部分被称为外奥尔特云，它才是球对称的。距估计，奥尔特云中约有几万亿颗直径在一千米以上的彗星，其总质量约为地球质量的几倍到几十倍。由于数量众多，在一些科普示意图中奥尔特云被画得像一个真正的云团一样，但事实上，奥尔特云中两个相邻小天体之间的平均距离约有几千万千米，是太阳系中天体分布最为稀疏的区域之一。

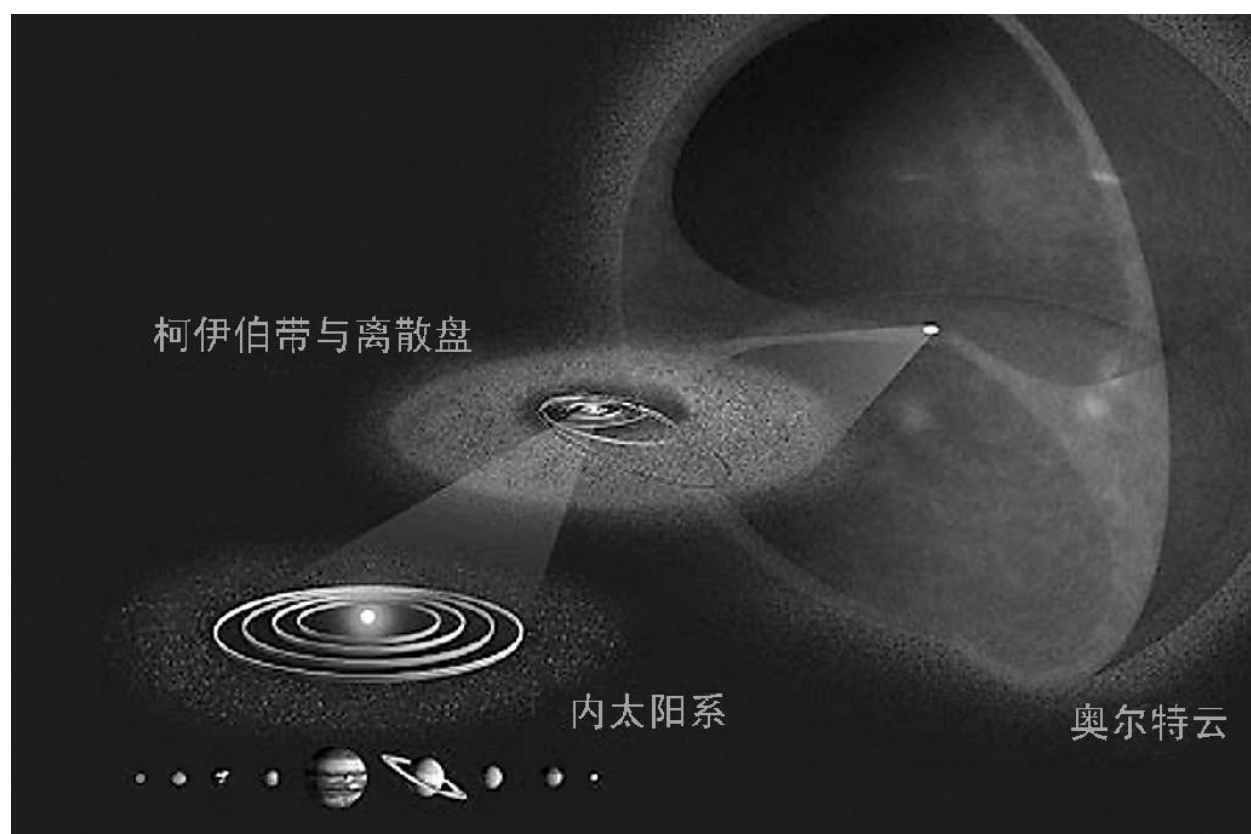
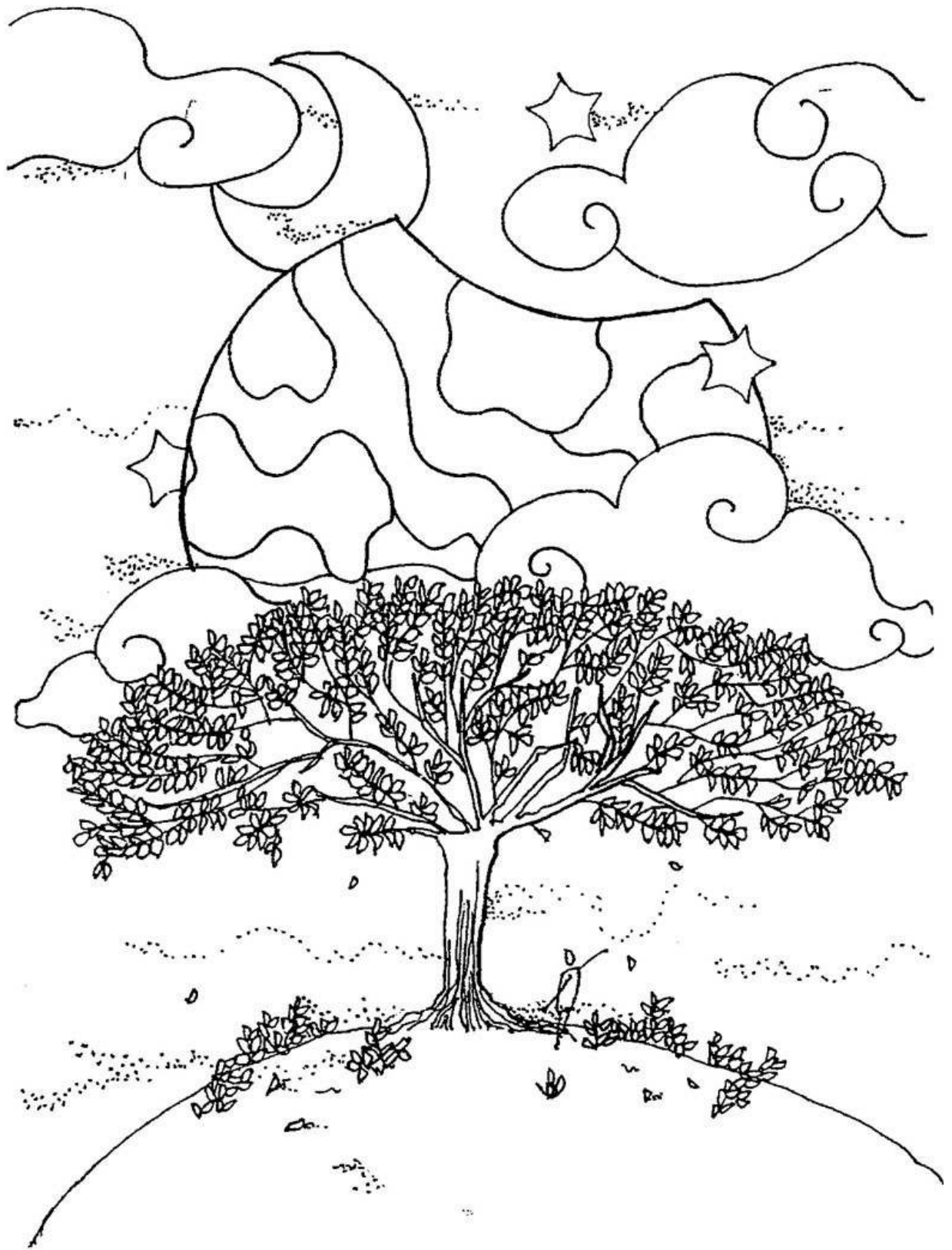


图17 奥尔特云及太阳系结构示意图



在距太阳如此遥远的地方为何会有这样一个奥尔特云呢？一些天文学家认为，与离散盘类似，奥尔特云最初是不存在的，如今构成奥尔特云的那些小天体最初与行星一样，形成于距太阳近得多的地方，后来是被外行星的引力作用甩了出去，才形成了奥尔特云。奥尔特云中的小天体由于距太阳极其遥远，很容易受银河系引力场的潮汐作用及附近恒星引力场的干扰，那些干扰会使得其中一部分小天体进入内太阳系，从而成为长周期彗星。

奥尔特云距我们如此遥远，而且包含的又大都是小天体，读者们也许会以为除直接来自那里的长周期彗星外，我们不太可能观测到任何属于奥尔特云的天体。其实不然。这倒不是因为我们有能力观测到几千乃至几万天文单位之外的小天体，而是因为奥尔特云并不是一个界限分明的区域。少数奥尔特云天体的轨道离我们相当近，甚至能近到可被直接观测到的程度。2003年，美国帕洛马天文台（Palomar Observatory）的天文学家布朗（Michael Brown，他也是创神星的发现者之一）发现了一个临时编号为2003VB₁₂（正式编号为90377）的海外天体，它的轨道远日点距离约为976天文单位，近日点距离也有76天文单位。这个天体的块头很大（否则就不会被发现了），直径约有1500千米，曾一度被当成第十大行星的候选者（当时阋神星尚未被发现）。天文学家们给它取了一个专门的名称：赛德娜（Sedna，因纽特神话中的海洋生物之神）。一般认为，赛德娜是属于内奥尔特云的天体^{[\[2\]](#)}。除赛德娜外，还有一个我们非常熟悉，有些读者甚至用肉眼都曾看到过的天体——哈雷彗星——也被认为是有可能来自奥尔特云的。哈雷彗星虽然是一颗短周期彗星，但很多天文学家认为，它是从奥尔特云进入巨行星的引力范围后受后者的干扰才成为短周期彗星的。

奥尔特云究竟有多大呢？今天的很多天文学家认为它的范围延伸到距太阳约50000天文单位的地方，但也有人像奥尔特当年一样，认为它

延伸得更远，直到太阳引力控制范围的最边缘。这一边缘大约在距太阳100000~200000天文单位处，在那之外，银河系引力场的潮汐作用及附近恒星的引力作用将超过太阳的引力。（请读者想一想，我们为什么在提到银河系引力场时强调“潮汐作用”，而在提到附近恒星的引力场时不强调这一点？）如果那样的话，奥尔特云的外边缘应该就是太阳系的疆界了。

不过，奥尔特云未必是太阳系疆界附近的唯一秘密。1984年，美国芝加哥大学的古生物学家劳普（David Raup）和塞普考斯基（Jack Sepkoski）在对过去两亿五千万年间地球上的大规模生物灭绝状况进行研究后提出，那种灭绝似乎平均每隔2600万年发生一次，而且有迹象表明其中至少有两次似乎与大陨星撞击地球的时间相吻合（其中最著名的一次被认为是发生在距今约6500万年的白垩纪末期，导致包括恐龙在内的大量生物灭绝）。同年，美国加州大学的物理学家马勒（Richard Muller）等人提出了一个大胆的猜测，认为太阳可能有一颗游弋在太阳系边缘的伴星，这颗伴星是一颗褐矮星或红矮星（褐矮星的质量约在木星质量的13~75倍之间，红矮星的质量约在木星质量的75~500倍之间），它距太阳最远时约有2.4光年（感兴趣的读者请根据上下文提供的信息，计算一下它离太阳最近时的距离）。这颗伴星每隔2600万年经过奥尔特云的一部分，在它的引力干扰下，大量的奥尔特云天体会脱离原先的轨道而进入内太阳系，其中个别天体会与地球相撞，从而造成大规模的生物灭绝。由于这颗伴星所起的可怕作用，它被称为内梅西斯（Nemesis），这是希腊神话中的复仇女神。如果太阳真的有这样一颗伴星，并且它真的有如人们所猜测的那种作用，那它无疑是太阳系疆界附近最可怕的天体⁽³⁾。即便如此，我们也不必害怕，因为按照那些科学家的说法，地球上最近一次大规模生物灭绝大约发生在距今五百万年以前，那么下一次同类事件——如果有的话——就该是两千多万年之后的事了。那时假如人类还存在，想必该有足够的智慧来避免灾难。

我们有关太阳系疆界的故事在这里就要与读者说再见了，但人类探索太阳系疆界的事业却远未结束，这样的事业有一个美丽的名字叫科学，她值得人们去做永生的探索。

注释

[\[1\]](#) 奥尔特并不是最早提出彗星大本营概念的天文学家。1932年，爱沙尼亚天文学家欧皮克（Ernst Öpik）曾经提出过彗星来自太阳系边缘的一片“云”的假设。此外，早年曾有一些天文学家认为短周期彗星也来自奥尔特云，只不过是在接近内太阳系时受到巨行星的影响而被俘获成了短周期彗星。但具体的计算及模拟表明，小天体从遥远的奥尔特云进入并被俘获在内太阳系的概率非常小，不足以解释观测到的短周期彗星的数量。而且来自奥尔特云的新彗星的轨道倾角分布也与短周期彗星的倾角分布有着显著差异。因此后来人们放弃了这一假设（但个别短周期彗星——比如哈雷彗星——仍被认为是有可能来自奥尔特云）。

[\[2\]](#) 2000年，罗威尔天文台发现的一个临时编号为2000CR₁₀₅（正式编号为148209），远日点距离约394天文单位，近日点距离约44天文单位的小天体也被认为有可能属于内奥尔特云，但争议较大。

[\[3\]](#) 需要提醒读者注意的是，有关太阳伴星的猜测目前只有很少的支持者，其学术地位远低于有关奥尔特云的猜测。

附录 冥王星沉浮记^{[\(1\)](#)}

引言

如果你徜徉在纽约曼哈顿的街头，也许会被一座特别的雕像所吸引，那便是矗立在著名地标性建筑洛克菲勒中心（Rockefeller Center）前的阿特拉斯（Atlas）雕像（图18）。阿特拉斯是希腊神话中象征着力量与坚忍的巨神，在他的肩上扛着整个天球^{[\(2\)](#)}。如果有办法细看的话，你也许还会惊讶地发现，这座雕像的天球之上只刻着八颗行星。比雕像的落成早七年就已发现，直到2006年才被降级的冥王星竟然不在其中。是艺术家未卜先知吗？不是。原来，这座落成于1937年的雕像是20世纪20年代设计的，当时冥王星尚未被发现，天球上自然也就没它的位置了。不过，那原本已成为缺陷的八大行星在相隔大半个世纪之后重新变得贴切，是谁也不曾料到的。



图18 洛克菲勒中心前的阿特拉斯雕像（弧形的天球支架上刻有行星符号）

如今距离冥王星的降级已时隔多年，冥王星是如何一步步走向降级的？降级后人们的反应又如何呢？在本附录中，我们将依照时间的顺序来回顾一下这颗昔日行星的“命运”沉浮。

冥王星降级倒计时**76**年

冥王星是1930年由美国罗威尔天文台的一位当时仅仅是观测助理的年轻天文学家汤博（Clyde Tombaugh）发现的（它因此而被称为“美国行星”）。与其他八大行星不同的是，冥王星的行星地位受到过多次怀疑。在发现之初，它曾被视为是一颗被理论所预言的新行星。但人们很快就发现，无论它的质量还是轨道，都与理论预言存在较大的差异。因此早在它被发现的那一年，就有人因其与理论预言不相符合，而怀疑它并非太阳系的第九大行星。不过那种怀疑并不成立，因为当时有关新行星的预言是错误的^{[\(3\)](#)}，与错误的预言不相吻合是不能怪冥王星的。

冥王星降级倒计时**50**年

1956年2月，冥王星的行星地位再次遭到了怀疑，美籍荷兰裔天文学家柯伊伯（Gerard Kuiper）在接受美国《时代》周刊的采访时表示，冥王星的自转周期超过6天，对于行星来说显得太慢了。柯伊伯是一位著名的行星天文学家，以他名字命名的柯伊伯带将在50年后成为冥王星降级的真正原因，但他以自转太慢为由怀疑冥王星的行星地位，却是站不住脚的。我们现在知道，水星的自转周期约为59天，金星的更是长达243天，都比冥王星转得更慢^{[\(4\)](#)}。

冥王星降级倒计时**28**年

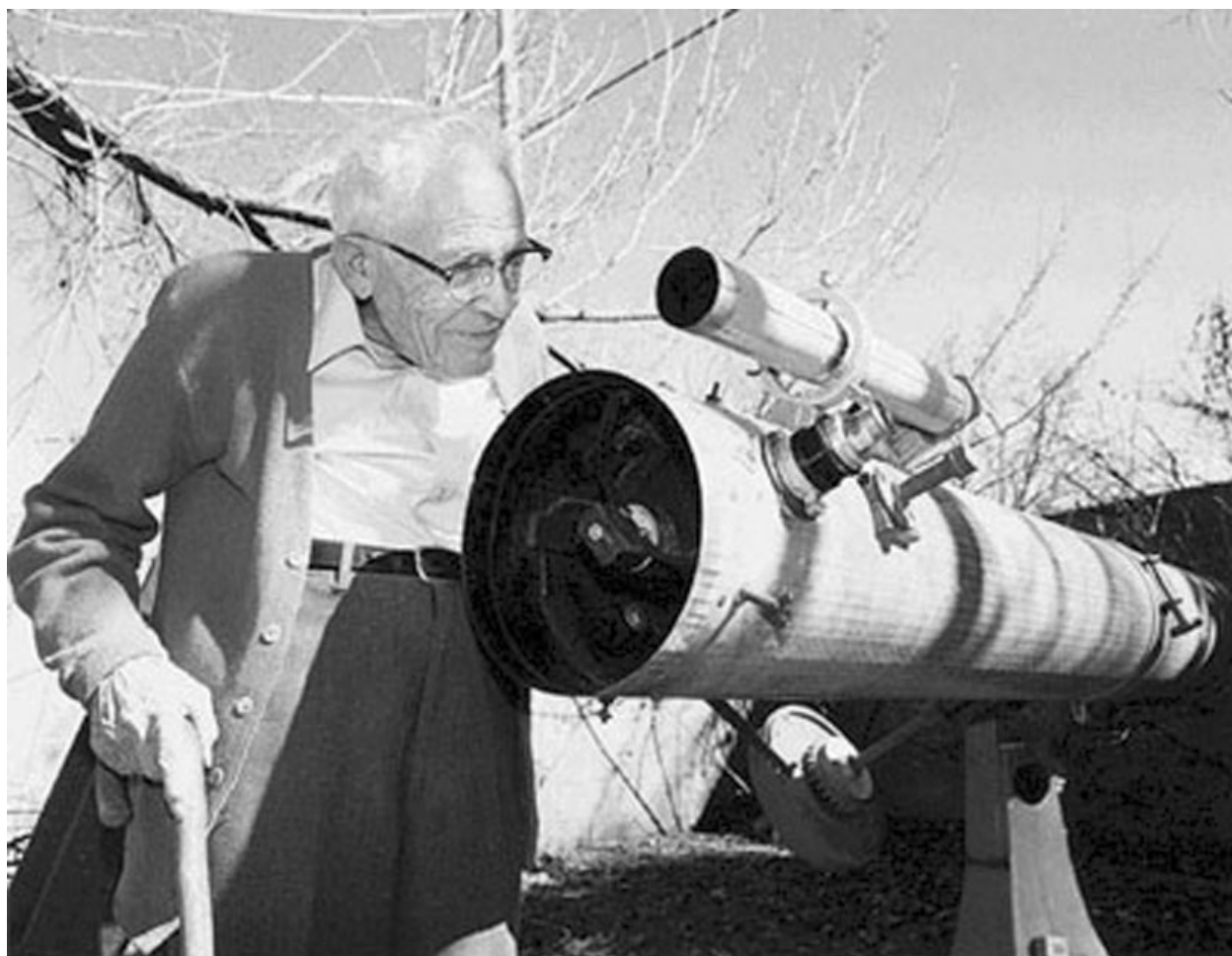
1978年6月，冥王星迎来了一个对其行星地位来说喜忧参半的消

息：它的卫星卡戎（Charon）被美国海军天文台的天文学家所发现。通过观测卡戎的运动，冥王星的质量首次得到了较为精确的测定，结果竟然还不到月球质量的1/5，这显然是个坏消息。但另一方面，很多天文学家相信，卫星按定义就是绕行星运转的天体，冥王星既然有卫星，它自己当然就只能行星了，因此这同时又是一个好消息。不过这好消息背后的理据在1994年遭到了破灭。那一年，天文学家们发现了小行星的卫星⁽⁵⁾，从而使得拥有卫星不再是行星的专利。

冥王星降级倒计时14年

1992年，冥王星作为太阳系中海王星以外之唯一天体（彗星不算）的地位宣告不保。自1992年起，人们在海王星之外陆续发现了越来越多的新天体（统称为海外天体），它们的大小虽暂时还不能与冥王星相比，但它们的出现越来越证实了天文学家们早在20世纪中叶就提出过的一种观点，即海王星之外存在大量小天体，它们都是行星演化的“半成品”，冥王星有可能是它们中的一员⁽⁶⁾。

“山雨欲来风满楼”，至此，冥王星的“命运”已岌岌可危，这危机惊动了一个人，他就是昔日那位罗威尔天文台的年轻观测助理，如今已德高望重的冥王星发现者汤博。



去世前不久的汤博

冥王星降级倒计时**12**年

1994年12月，已经88岁高龄的汤博给美国科普杂志《天空与望远镜》（Sky & Telescope）写了一封信，为冥王星的“命运”做最后一搏。在信中他主张像维持其他天文命名体系——比如恒星的光谱命名及星座的命名那样保留冥王星的行星地位。他并且主张以17等星作为分界，将新近发现的海王星以外暗于17等的小天体命名为柯伊伯小天体

（Kuiperoids），以区别于冥王星。可惜的是，这些主张都没什么说服力，以17等星（而且还是视星等）为分界更是充满了随意性^{[\(7\)](#)}。两年

之后，汤博离开了人世。

冥王星降级倒计时**6**年

2000年2月，位于纽约曼哈顿的海登天文馆（Hayden Planetarium）作出了一个大胆的决定，在太阳系的行星模型中破天荒地去掉了冥王星。2001年1月22日，这一公开的秘密被《纽约时报》的记者所发现，并以《冥王星不是行星吗？只在纽约》（Pluto's Not a Planet? Only in New York）为题在头版作了报道。那一天，天文馆主任泰森（Neil Tyson）的电话留言及电子邮箱均被雪片般飞来的询问与质疑挤爆。不过当时学术界有关冥王星行星地位的意见已足够分歧，泰森成功地顶住了压力。

冥王星降级倒计时**1**年

2005年1月，美国天文学家布朗（Michael Brown）在检查旧的观测相片时发现了一颗比冥王星更大的海外天体：阋神星。从此冥王星不再只是海王星以外的唯一天体（彗星不算），甚至连最大的天体也不再是了。这一发现在很大程度上成为了“压垮”冥王星行星地位的最后一根稻草。

冥王星降级倒计时**7**个月

2006年1月，英国广播公司（BBC）采访了地球上最后一位与冥王星有直接渊源的人：英国退休女教师费尔（Venetia Phair）。76年前，年仅11岁的她提议了冥王星的名字^{[\(8\)](#)}。不过，在被问及冥王星的“命运”危机有何看法时，费尔表示自己年事已高，不再关心此事，但她乐意看到冥王星继续当行星。三个月后，费尔也离开了人世。

冥王星降级零时

2006年8月24日，冥王星“命运”水落石出的时刻终于来临。国际天文联合会（IAU）的424位天文学家在捷克共和国的首都布拉格（Prague）就行星的定义及冥王星的地位问题举行投票。在投票结果即将宣布的那一刻，无数记者在场外屏息等候，用海登天文馆主任泰森的形容，那满场的寂静宛如梵蒂冈教廷任命新教宗前，教徒们在宫殿外屏息等候的情形。一个科学事件引起如此关注是不多见的。

投票的结果是：冥王星降级成了矮行星（dwarf planet）。

一石激起千层浪！

在那一刻之前，也许很少有人真正关心过那个远在60亿公里之外的由岩石与寒冰组成的遥远球体，有关冥王星“命运”的争议也基本局限在科学界之内；在那一刻之后，整个事件被骤然披上了浓厚的文化色彩。学生、政客、占星师、宗教信徒、科学爱好者等，全都加入了关注行列，并发表了种种意见。如果说此前的倒计时所记录的主要是科学事件，那么此后的时钟却记录了很多文化及社会事件。

冥王星降级后几分钟

冥王星降级的消息立刻在全球媒体上占据了重要版面，有人甚至精心杜撰了许多搞笑标题，比如《美国联邦经费不足导致太阳系裁员》，《民主党人拒绝向冥王星提供援助》，《冥王星降级违宪》，等等^[9]。不过真正搞笑的要数美国加州议会的一份真实的抗议提案。那份提案事先就已拟定，并在国际天文联合会投票结束几分钟之后就提了出来。加州之所以如此有备而来，是因为“冥王星”这一名称在加州有另外一层含义，它是总部位于加州的迪斯尼乐园（Disneyland）中一条深受

孩子们喜爱的宠物狗的名字。加州议会在提案中郑重表示，冥王星的降级将会“伤害加州的孩子”[\(10\)](#)。不过，议员们的热心并未得到宠物狗的真正主人——迪斯尼公司——的响应。迪斯尼公司表示，宠物狗“冥王星”除了偶尔会对着月亮嚎叫几声外，对其他天体并无兴趣。加州议会的冥王星提案最终搁浅。

冥王星降级后4个月

冥王星的降级也让很多冥王星爱好者感到不满，他们以各种方式表达了对冥王星的深切怀念，其中包括组建冥王星粉丝团，制作小宣传品，开办请愿网站，等等。2006年底，美国方言协会（American Dialect Society）宣布将冥王星Pluto由名词提升为动词，用法为to pluto或to be plutoed。仿照当下中文网上颇为流行的“被”字短语（比如“被就业”、“被增长”等），该动词可译为“被冥”，其含义为“像冥王星一样被贬”。冥王星在天文学上被贬，却在词义上获得了提升，算是略有补偿吧。

冥王星降级后1~3年

自冥王星降级后的第二年起，美国的另两个州也步加州后尘提出了冥王星提案，并且还得到了通过。那两个州都与冥王星的发现者汤博有着密切关系。其中一个是新墨西哥州，那是汤博后半生的居住地，也是他任教18年之久的新墨西哥州立大学（New Mexico State University）的所在地。2007年3月8日，新墨西哥州议会通过决议，宣布冥王星在该州仍然是行星。另一个是伊利诺伊州，那是汤博的出生地。2009年2月26日，伊利诺伊州议会也通过决议，宣布冥王星在该州为行星。因此，现在我们可以仿照《纽约时报》当年的标题说一句：“冥王星是行星吗？”

只在新墨西哥州和伊利诺伊州”。与汤博有关的另一个重要地点——发现冥王星的罗威尔天文台——也不落后，别出心裁地在捐款箱上设计了几个小小的选项，让大家用钱包来投票，结果——如所预料的——是支持冥王星为行星的参观者为数最多（图19）。



图19 罗威尔天文台的捐款箱

另一方面，天文学家们的意见也并非铁板一块。冥王星虽然被降级了，许多天文学家对它的“爱心”却依然不改。冥王星“被冥”后不久，美

国行星科学研究所（Planetary Science Institute）的主任赛克斯（Mark Sykes）就牵头发表了一份由304位科学家签名的请愿书，宣布不承认国际天文联合会的投票结果。304这一人数大有直逼国际天文联合会的投票人数424之势，不过简单的统计表明，签名者中绝大多数是美国科学家，非美国的只有不到20人（而国际天文联合会中的非美国科学家占2/3）。看来对冥王星地位的看法即便在学术界之内也不是单纯的学术问题⁽¹¹⁾。国际天文联合会收到的抗议信也有着同样鲜明的国别色彩，超过90%是来自美国民众的，这与冥王星是“美国行星”显然不无关系。

常言道：解铃还需系铃人。学术问题归根到底还是要用学术手段来解决。2008年8月，一百多位天文学家聚集在美国的约翰·霍普金斯大学（John Hopkins University），再次就行星定义展开了讨论。在讨论中，很多天文学家表达了自己的看法。那些看法从支持国际天文联合会的定义，到将行星俱乐部扩招几十倍⁽¹²⁾；从以保护“文化遗产”为名保留冥王星的“行星籍”，到干脆将月球也升级为行星，林林总总，应有尽有。由于分歧实在太太大，后来的会议简报只列出了一条不无搞笑意味的共识，叫做“没有共识的共识”（agree to disagree）。

在针对国际天文联合会有关冥王星地位所作的表决的全部质疑中，最有技术含量的理由是参与表决的人数太少，还不到全体会员人数的5%，从而缺乏代表性。这一理由听起来不无道理，因而被许多人所支持，但它其实并非真的很有力，因为当时的表决结果是以90%的大比数通过的，远大于统计误差，很难被单纯的人数增加所改变。不过另一方面，这些年来天文学家们始终无法就行星定义达成共识这一事实，从一个侧面显示出当年的表决确有值得商榷之处，只不过这商榷之处恐怕不是人数太少，而是在于选项太少，即在表决时只有一份提案可供选择，从而无可避免地带有片面性。这就好比在晚饭时间，让一群人选择吃川菜还是不吃，多数人——包括不太喜欢川菜的人——都会选择吃川菜；

但如果选项增加为：吃川菜、粤菜、鲁菜、浙菜还是不吃，意见也许就会相当分歧。

尾声

有关冥王星这颗“美国行星”的争议看来还将持续很长时间。虽然有那么多人在关注，我们对冥王星的真正了解却少之又少，甚至连一张像样的图片都拿不出来。为了改变这一局面，2006年1月19日，美国国家航空航天局发射了人类有史以来第一个冥王星探测器：新视界（New Horizons）。这个探测器上除了观测仪器外，还携带着冥王星发现者汤博的部分骨灰，这位来自伊利诺伊州的“农民的儿子”将在2015年魂游自己所发现的冥王星。当他出发时，冥王星还是一颗行星，如今它却只是一颗编号为134340的矮行星了。不过，让我们且把名份之争放在一边，翘首期待“新视界”探测器掠过冥王星的那一刻吧，无论我们如何称呼冥王星，那都将是一个激动人心的时刻（图20）。



图20 “新视界”探测器飞临冥王星的想像图

注释

[\[1\]](#) 本文的删节版曾发表于2009年10月的《科学画报》。

[\[2\]](#) 阿特拉斯（Atlas）还是英文单词atlas（地图册）的词源。

[\[3\]](#) 有关这一点的详细介绍，请参阅第23、24、27等章。

[\[4\]](#) 读者也许会觉得奇怪，像柯伊伯那样的天文学家怎么会把像自转速度那样细枝末节的性质作为怀疑冥王星行星地位的理由？其实他的真正理由是：像冥王星那样慢的自转当时只在卫星中被发现过，因此冥王星的缓慢自转说明它有可能是一颗侥幸逃脱海王星引力束缚的卫星。这种将巨行星的某些卫星与像冥王星那样的柯伊伯带天体联系起来的观点是颇有远见的。虽然我们现在并不认为冥王星是逃脱海王星引力束缚的卫星，但相反的过程，即柯伊伯带天体

被俘获成为海王星（或其他巨行星）卫星的过程却得到了不少天文学家的认同，比如海王星的卫星Triton（海卫一）就被认为很可能是遭俘获的柯伊伯带天体。

[\[5\]](#) 这一发现是通过美国的“伽利略”号探测器得到的，所发现的是围绕小行星Ida（艾达）运转的卫星。小行星Ida是一个形状不规则的天体，平均线度为31.4公里，它的卫星Dactyls（戴克泰）的平均线度则为1.4公里。

[\[6\]](#) 有关这一观点的详细介绍，请参阅第28章。

[\[7\]](#) 读者也许会问：汤博为什么选17等星这样一个特殊星等？答案很简单：那是汤博自己曾经搜索过的最暗天体的视星等。以自己的天文搜索能力作为天体分界的标准，在学术上显然是没有任何说服力的。

[\[8\]](#) 有关这一点，请参阅第26章。

[\[9\]](#) 这些搞笑标题来自美国的一份政治幽默杂志The People's Cube。

[\[10\]](#) 加州议案的抗议理由还包括冥王星的降级会“损害某些担忧普适常数（注：指行星数目）稳定性的加州人的心理健康”及“扩大加州的财政赤字”。

[\[11\]](#) 科学家也是人，他们在考虑问题——尤其是像冥王星身份这种介于主观与客观之间的问题——时也不可避免地会掺入个人情感甚至民族情感。从民族情感上讲，美国民众（包括科学家）希望冥王星保留行星身份者为数较多，其他国家的人则大都无所谓；从个人情感上讲，冥王星的发现者汤博、命名者费尔、冥王星探测计划的主管者斯特恩（Alan Stern）等都主张保留冥王星的行星身份，而柯伊伯带的提出者柯伊伯与发现者朱惠特、卢简等则持相反看法。

[\[12\]](#) 扩招几十倍的方法是放弃国际天文联合会的定义中“扫清自己轨道附近的区域”这一条件，这样一来潜在的行星数目有可能增加到几百甚至更高（关于这一点，请参阅第31章）。

术语表

矮行星（**dwarf planet**）

矮行星是国际天文联合会于2006年8月24日结合行星新定义而提出的太阳系天体的新类别，太阳系内的矮行星是同时满足以下四个条件的天体：（1）围绕太阳公转；（2）具有足够的质量使自身引力克服刚体应力，从而具有（近球形的）流体静力平衡形状；（3）没有扫清自己轨道附近的区域；（4）不是卫星。截至2009年8月，太阳系中共有五个天体被定为矮行星，它们分别是：谷神星（Ceres）、冥王星（Pluto）、阋神星（Eris）、鸟神星（Makemake）和妊神星（Haumea）。这一数目今后无疑将会增加。

奥尔特云（**Oort cloud**）

奥尔特云是以荷兰天文学家奥尔特（Jan Oort）的名字命名的假想中的长周期彗星大本营，其范围有可能一直延伸到太阳引力控制范围的最边缘（距太阳100000~200000天文单位）。奥尔特云有可能存在内外之分，距太阳20000天文单位以内的内奥尔特云——也叫希尔云（Hill cloud）——呈圆环形分布，在那之外的外奥尔特云则呈球对称分布。据估计，奥尔特云中约有几万亿个直径在一千米以上的天体。奥尔特云天体距太阳的平均距离虽然极远，但个别天体的近日点距离却有可能并不太大，从而能被观测到，比如今后有可能会被提升为矮行星的太阳系小天体赛德娜（Sedna），就有可能是一个奥尔特云天体。

表观逆行（**apparent retrograde motion**）

观测天文学上的表观逆行，是指因地球（或观测者所在的其他参照系）本身的运动而造成的被观测天体相对于背景星空的表观运动与其相对于太阳的真实运动相反的现象。从某种意义上讲，如果我们相信行星的运动受简单规律所引导，那么表观逆行可以认为地球本身也是行星，从而也在运动的很有力的证据之一。不过在早年的历史上，人们宁愿用包含大量本轮、均轮的复杂模型来解释包括表观逆行在内的行星运动，也不愿轻易接受地球也在运动的观念。

电荷耦合器件（**charge coupled device**）

电荷耦合器件（简称CCD）是一种能够传输及存储电荷的半导体器件，它的一项很重要的用途是与光电器件相结合，制成可以取代传统胶片的感光器件。CCD是美国贝尔实验室（Bell Labs）的科学家博伊尔（Willard Boyle）和史密斯（George Smith）于1969年发明的（博伊尔和史密斯因此而获得了2009年的诺贝尔物理学奖），它已成为现代数码影像技术及观测天文学中不可或缺的工具。CCD作为感光器件的最大优点之一是具有极高的敏感度，能对70%甚至更大比例的入射光作出反应（普通照相胶片的这一比例还不到10%）。另外，CCD所具有的影像记录数字化的特点，还为计算机处理提供了极大的便利。在历史上，柯伊伯带天体的发现就借助了CCD的帮助。

广义相对论（**general theory of relativity**）

广义相对论是物理学家爱因斯坦（Albert Einstein）于1915年底提出的引力理论。广义相对论将引力效应归结为时空的弯曲，是物理理论几何化的一个范例。自提出以来，广义相对论的各种预言已得到了大量观测及实验的支持，直到今天仍是描述万有引力的最佳理论。广义相对论

不仅是现代宇宙学及强引力场研究的基础，而且也是对弱引力场下的精密效应进行分析的重要工具，它的影响甚至包括了诸如全球卫星定位系统这样的应用领域。

国际天文联合会（**International Astronomical Union**）

国际天文联合会是一个由职业天文学家组成的国际机构，成立于1919年，总部位于法国的巴黎。国际天文联合会目前共有一万多名会员，分布于近百个不同的国家。国际天文联合会的主要职责包括组织国际天文会议，对天体及天体表面地貌进行命名等。国际天文联合会近期最具争议的一个举动是于2006年8月24日投票通过了有关太阳系行星的定义，并将76年来一直被视为行星的冥王星分类为了矮行星。

海王星档案（**Neptune files**）

海王星档案是一批与海王星发现有关的历史文件，主要包括海王星发现前后英国天文学家艾里（**George Airy**）与国内外同行的通信及其他资料。海王星档案最初被艾里存放于格林威治天文台，但在20世纪中期被恒星天文学家艾根秘密“借”走，直到艾根去世后的1998年才重见天日。海王星档案的部分内容目前已在互联网上公布。个别历史学者曾依据海王星档案对传统的海王星发现史提出了质疑，但那些质疑带有较强的阴谋论色彩，迄今并无足够的说服力成为史学界的主流观点。

彗星（**comet**）

彗星一词的希腊文原意是“头发”（后来被亚里斯多德引申为“带头发的星星”），是围绕太阳运动的太阳系小天体的一种。在接近太阳时，彗星上的挥发性物质会在太阳辐射及太阳风的作用下形成长长的彗

尾（“带头发的星星”之名由此而来）。彗星是天空中除行星外最常见的移动天体，历史上天文学家们曾多次将新发现的行星或小行星误当成彗星。太阳系内的彗星按轨道周期可大致分为两类：周期在200年以下的称为短周期彗星，它们大都来自柯伊伯带及离散盘；周期在200年以上的称为长周期彗星，它们被认为是来自奥尔特云。

〔角〕秒（**arc second**）

〔角〕秒是观测天文学上常用的角度单位，1〔角〕秒等于1〔角〕分的1/60，或1度的1/3600，或圆周（360度）的1/1296000。肉眼观测所能达到的最高精度通常为几十〔角〕秒。

康德-拉普拉斯星云假说（**Kant-Laplace nebular hypothesis**）

康德-拉普拉斯星云假说是有关太阳系起源的假说，最初的想法是由瑞典科学家斯韦登伯格（Emanuel Swedenborg）于1734年提出的。1755年，德国哲学家康德（Immanuel Kant）发展了这一想法。1796年，法国数学家拉普拉斯（Pierre-Simon Laplace）也独立地提出了类似的假说。康德-拉普拉斯星云假说认为太阳系是由一团星际尘埃云收缩凝聚而成的，这一想法成为了目前太阳系（以及其他行星系统）演化学说中的主流想法。

柯伊伯带（**Kuiper belt**）

柯伊伯带也称为埃奇沃斯-柯伊伯带，是20世纪中叶先后由包括爱尔兰天文学家埃奇沃斯（Kenneth Edgeworth）和美籍荷兰裔天文学家柯伊伯（Gerard Kuiper）在内的多位天文学家从理论上提出，并在20世纪末得到观测证实的天体带。柯伊伯带与太阳的距离约为30～55天文单

位。一般认为，柯伊伯带天体是行星演化过程中的半成品。据估计，柯伊伯带中仅直径大于100千米的天体就有70000个以上，其中最著名（并且也最大）的是矮行星冥王星。柯伊伯带与离散盘被认为是太阳系中短周期彗星的大本营。

离散盘（**scattered disc**）

离散盘是太阳系外围的一个盘状区域，与太阳的距离从30~35天文单位延伸到100天文单位甚至更远。离散盘天体的轨道通常具有较大的椭率，半长径通常在50天文单位以上，其中最著名（迄今所知也最大）的天体是矮行星阋神星（Eris）。目前天文学家们对离散盘的了解还很有限，一般认为，离散盘中的天体有可能是被外行星的引力甩出来的柯伊伯带天体。

闪视比较仪（**blink comparator**）

闪视比较仪是通过快速切换的方法来对比两张不同相片的仪器。闪视比较仪特别适合于寻找在两次拍摄间亮度或位置发生变化的天体。在历史上，冥王星就是通过闪视比较仪发现的。随着电荷耦合器件及计算机图像对比与处理技术的普及，闪视比较仪的重要性已有了显著的下降。

视星等（**apparent magnitude**）

视星等是扣除了大气层的影响后，天体相对于地面观测者的表观亮度。视星等采用的是对数标度，其中正常肉眼所能看见的最暗天体定义为6等，比这一天体亮100倍的天体定义为1等（因此视星等每相差1等，亮度相差 $100^{1/5} \approx 2.512$ 倍）。观测天文学上的一些典型的视星等为：太

阳-26.73，满月-12.6，最亮时的金星-4.6，最亮时的天王星5.5，最亮时的谷神星6.7，最亮时的冥王星13.6，口径8米的地面光学望远镜所能观测的最暗天体的视星等为27，哈勃望远镜所能观测的最暗天体的视星等为30。

提丢斯—波德定则（**Titius-Bode law**）

提丢斯—波德定则是德国天文学家提丢斯（Johann Titius）于1766年提出的太阳系天体分布经验规律。按照这一定则，太阳系各行星的轨道半径（以地球轨道半径为单位） r_n 满足 $r_n = 0.4 + 0.3 \times 2^n$ （其中水星对应于 $n = -\infty$ ，其余行星及小行星带自内向外依次对应于 $n = 0, 1, 2, 3$ 等）。这一定则经过德国天文学家波德（Johann Bode）的“借用”及传播后广为人知，并在小行星带的发现及海王星的轨道计算中起到过一定作用。提丢斯—波德定则对于海王星以内的各行星及小行星是不错的近似，在那之外则基本无效。一般认为，提丢斯—波德定则并无理论依据，有可能是轨道共振及初始条件的共同结果，也可能只是巧合。

天体力学（**celestial mechanics**）

天体力学是运用力学原理研究天体运动的天文学分支。天体力学通常用于计算已知天体（包括人造天体）的运动，但在历史上也曾被用于推算未知天体的位置，其中最成功的例子是对海王星位置的预言。天体力学中的一些著名问题——比如三体问题——曾引起数学家与物理学家的强烈兴趣及深入研究。在精密的天体力学计算中有时需要引进相对论修正，其中最著名的例子是在水星近日点进动的计算中引进广义相对论修正。

天文单位（**astronomical unit**）

天文单位是行星天文学上最常用的距离计量单位，它近似等于地球与太阳的平均距离，或1.496亿千米。它在国际单位制中的严格定义为：在太阳引力作用下沿圆轨道以每天0.01720209895弧度的角速度运动的试验粒子的轨道半径。严格地讲，天文单位的大小是不恒定的。

（感兴趣的读者请思考一下，哪些因素会导致上述定义下的天文单位不恒定。）

牛顿万有引力定律（**Newton's law of universal gravitation**）

牛顿万有引力定律是描述有质量物体之间引力相互作用的物理学定律，它是英国物理学家牛顿（Isaac Newton）在1687年出版的著作《自然哲学的数学原理》中发表的（他的一些同时代人也有过类似的想法）。按照牛顿万有引力定律，两个线度可以忽略的有质量物体之间的引力的大小正比于两个物体质量的乘积，平方反比于两个物体的距离，方向则沿两个物体的连线。牛顿万有引力定律在很长的时间里一直是天体力学的基础，并且直到今天依然适用于引力场不太强，运动速度不太快，对精度要求不太高的天体力学计算。

小行星带（**asteroid belt**）

小行星带是大致位于火星与木星轨道之间的环状分布的小天体群。小行星带中最早被发现的若干成员曾一度被误当成行星。按照目前人们对太阳系天体的分类，小行星带中最著名（并且也最大）的天体是矮行星谷神星（Ceres），其余按目前的分类则全都是太阳系小天体。据统计，小行星带中约有超过一百万个直径一千米以上的天体。

行星（**planet**）

行星一词的希腊文原意是“漫游者”，最初指的是太阳系内的金、木、水、火、土五大行星，在日心说被采纳后又增加了地球。在约定成俗几千年之后，国际天文联合会于2006年8月24日对太阳系内的行星进行了定义。按照这一定义，太阳系内的行星是同时满足以下三个条件的天体：（1）围绕太阳公转；（2）具有足够的质量使自身引力克服刚体应力，从而具有（近球形的）流体静力平衡形状；（3）扫清了自己轨道附近的区域。目前太阳系中共有八个行星，它们分别是：水星、金星、地球、火星、木星、土星、天王星和海王星。

人名索引

阿伯特 (Abbott Lowell)
阿拉果 (François Arago)
埃奇沃斯 (Kenneth Edgeworth)
艾根 (Olin Eggen)
艾里 (George Biddell Airy)
爱德华 (Edward Pickering)
爱因斯坦 (Albert Einstein)
奥伯斯 (Heinrich Wilhelm Olbers)
奥尔特 (Jan Oort)
奥里安尼 (Barnaba Oriani)
巴比涅特 (Jacques Babinet)
巴格代迪 (Baghdady)
比尔 (Thomas Bill)
毕达哥拉斯 (Pythagoras)
波德 (Johann Elert Bode)
波瓦德 (Alexis Bouvard)
伯克哈特 (Johann Karl Burckhardt)
布莱德利 (James Bradley)
布朗 (Michael Brown)
布鲁克斯 (C. J. Brookes)
查利斯 (James Challis)
达兰伯利 (Jean Baptiste Joseph Delambre)
达雷斯特新 (Heinrich Louis d'Arrest)
达利特 (Gabriel Dallet)

第谷（Tycho Brahe）
杜林冈德斯（Vicomte du Ligondes）
恩克（Johann Franz Encke）
菲克斯米尔纳（Alexander Fixlmillner）
费尔（Venetia Phair）
费尔南德斯（Julio Fernández）
费恩曼（Richard Feynman）
弗格森（James Ferguson）
弗拉姆斯蒂德（John Flamsteed）
弗莱马力奥（Camille Flammarion）
福布斯（George Forbes）
伽勒（Johann Gottfried Galle）
伽利略（Galileo Galilei）
伽诺夫斯基（Alexander Garnowsky）
盖-吕萨克（Joseph Louis Gay-Lussac）
盖洛特（Jean Baptiste Gaillot）
高斯（Carl Friedrich Gauss）
格里戈尔（Theodore Grigull）
哈丁（Karl Harding）
赫西（Thomas Hussey）
赫歇耳（William Herschel）
惠普尔（Fred Whipple）
霍恩斯比（Thomas Hornsby）
卡洛琳（Caroline Herschel）
卡梅伦（Alastair Cameron）
开普勒（Johannes Kepler）
康德（Immanuel Kant）

柯伊伯 (Gerard Kuiper)
科勒斯特姆 (Nick Kollerstrom)
拉比诺维茨 (David Rabinowitz)
拉格朗日 (Joseph Louis Lagrange)
拉兰德 (Joseph Lalande)
拉莫尼亚 (Pierre Charles Le Monnier)
拉普拉斯 (Pierre Simon Laplace)
莱克塞尔 (Anders Johan Lexell)
莱兰德 (Michel Lalande)
莱沙鲍特 (Edmond Lescarbault)
兰伯特 (Johann Heinrich Lambert)
兰普朗德 (Carl Lampland)
劳 (Hans-Emil Lau)
劳普 (David Raup)
勒维耶 (Urbain Le Verrier)
利奥纳德 (Frederick C. Leonard)
利普歇 (Hans Lippershey)
卢简 (Jane Luu)
罗林斯 (Dennis Rawlins)
罗威尔 (Percival Lowell)
马勒 (Richard Muller)
马斯登 (Brian Marsden)
马斯克林 (Nevil Maskelyne)
梅尚 (Pierre Méchain)
梅耶 (Tobias Mayer)
尼古拉 (Friedrich Bernhard Nicolai)
牛顿 (Isaac Newton)

诺森伯兰公爵（Duke of Northumberland）

欧拉（Leonhard Euler）

欧皮克（Ernst Öpik）

泡利（Wolfgang Pauli）

皮考克（George Peacock）

皮克林（William Pickering）

比亚奇（Giuseppe Piazzi）

珀金斯（Adam Perkins）

普特南（Roger Putnam）

乔治（George Adams）

萨隆（Bochart de Saron）

塞普考斯基（Jack Sepkoski）

沙普利（Harlow Shapley）

舒马赫（Heinrich Schumacher）

斯莱弗（Vesto Slipher）

斯坦迪什（Erland Myles Standish, Jr）

斯特恩（Alan Stern）

斯威福特（Lewis Swift）

泰森（Neil Tyson）

汤博（Clyde Tombaugh）

特鲁吉罗（Chad Trujillo）

提丢斯（Johann Daniel Titius）

托德（David Todd）

瓦尔兹（Benjamin Valz）

沃夫（Christian von Wolff）

沃森（James Watson）

西伊（Thomas See）

小赫歇耳 (John Herschel)

辛德 (John Russell Hind)

约翰·亚当斯 (John Couch Adams)

扎克 (Franz Xaver von Zach)

朱惠特 (David Jewitt)

术语索引

1992QB₁

2003UB₃₁₃

埃里斯

矮行星

奥尔特云

巴黎天文台

柏林天文台

半人马小行星

北双子天文台

贝尔实验室

表观逆行

长周期彗星

创神星

地球

电荷耦合器件 (CCD)

短周期彗星

法国经度局

分光双星

复仇女神

格林威治天文台

谷神星

国际天文联合会

国家光学天文台

哈佛学院天文台

哈雷彗星
海军天文台
海外天体
海王星
海王星档案
褐矮星
红矮星
黄道面
彗星
婚神星
火神星
火星
基特峰国家天文台
剑桥天文台
角秒
金星
军团勋章
卡戎
考普雷奖
柯伊伯带
柯伊伯带天体
离散柯伊伯带天体
离散盘天体
罗威尔天文台
茂纳基雅天文台
美国艺术与科学学院
冥王星

木星
内奥尔特云
鸟神星
诺森伯兰望远镜
帕洛马天文台
七英尺望远镜
乔治星
妊神星
塞罗托洛洛天文台
赛德娜
闪视比较仪
视星等
水星
斯皮策太空望远镜
死神星
太阳伴星
提丢斯—波德定则
天王星
天王星出轨之谜
土星
外奥尔特云
阋神星
小天体中心
小行星
小行星带
行星
行星定义

英国皇家学会

英国科学进步协会

智神星

最小平方法

参考文献

1. Bartusiak M. Archives of the Universe [M]. New York: Vintage Books, 2004
2. Beatty J K, Petersen C C, Chaikin A. The New Solar System. London: Cambridge University Press, 1999
3. Brookes C J. On the Prediction of Neptune [J]. Celestial Mechanics, 1970,3: 67-80.
4. Davies J. Beyond Pluto [M]. London: Cambridge University Press, 2001.
5. Graney C M. On the Accuracy of Galileo's Observations [J]. Baltic Astronomy, 2007, 16 (3) : 443-449.
6. Hoskin M. Bode's Law and the Discovery of Ceres [J]. Astrophysics and Space Science Library, 1993,183: 35.
7. Hoyt W G. Planets X and Pluto [M]. Tucson: University of Arizona Press, 1980.
8. Kollerstrom N. An Hiatus in History: The British Claim for Neptune's Co-prediction, 1845-1846 [J]. Hist of Sci, 2006, 44 (3): 349-371.
9. Littmann M. Planets Beyond: Discovering the Outer Solar System [M]. New York: Dover Publications, Inc., 2004
10. Miner E D. Uranus: The Planet, Rings and Satellites [M]. Hoboken: John Wiley & Sons, 1998
11. Motz L, Weaver J H. The Story of Astronomy [M]. New York: Perseus Publishing, 1995
12. Price F W. The Planet Observer's Handbook [M]. London: Cambridge University Press, 1994

13. Standage T. The Neptune File [M]. New York: Walker Publishing Company, Inc., 2000.
14. Tyson N. The Pluto Files: The Rise and Fall of America's Favorite Planet [M]. New York: W. W. Norton & Company, Inc., 2009.
15. Weintraub D A. Is Pluto a Planet: A Historical Journey through the Solar System [M]. Princeton: Princeton University Press, 2006.

理解科学丛书·赵峥科普著作

INVISIBLE
STAR
The Black Holes
and the River of Time

看不见的星

黑洞与时间之河

赵峥◎著

当爱因斯坦讲“上帝不掷骰子”时，他错了。对黑洞的思索向人们提示，上帝不仅掷骰子，而且有时还把骰子掷到人们看不到的地方去了，那地方就是黑洞。

霍金

S. W. HAWKING

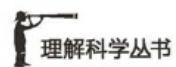
时间是什么，人不问我，我很清楚；一旦问起，我便茫然。

圣·奥古斯丁

ST. AUGUSTINUS

公元4世纪

清华大学出版社



INVISIBLE
STAR
The Black Hole,
and the River of Time

看不见的星

黑洞与时间之河

赵峥◎著

清华大学出版社
北京

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

看不见的星：黑洞与时间之河 / 赵峥著. —北京：清华大学出版社，
2014

（理解科学丛书）

ISBN 978-7-302-38589-9

I. ①看... II. ①赵... III. ①黑洞—研究 IV. ①P145.8

中国版本图书馆CIP数据核字（2014）第273649号

责任编辑：邹开颜

封面设计：蔡小波

责任校对：赵丽敏

责任印制：刘海龙

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社总机：010-62770175

邮购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市少明印务有限公司

经销：全国新华书店

开本：165mm×240mm

印张：13.25

字数：201千字

版次：2014年12月第1版

印次：2014年12月第1次印刷

产品编号：055376-01



引言

我第一次接触到黑洞这个概念，是在1958年前后。当时我在北京一中上初中，这所古老的中学有一个历史悠久的图书馆。我的同班好友裴申是一位天文爱好者，他从学校图书馆借到一本叫做《每月之星》的科普读物，听他谈论后我觉得非常有趣，于是也去图书馆借了一本。

这本书把中国古代的星座与希腊的星座对照介绍，不仅有美丽的神话，而且有当时最先进的科学知识。书中介绍了白矮星，那里有令人咋舌的地球人无法想象的高密度物质，体积像小酒杯大小的一块就重达1吨左右。书中还介绍了尚未发现的中子星，而且谈到爱因斯坦的广义相对论预言存在一种“看不见的星”。由于引力太大，这种星的光跑不出来，我们看不见。这种“看不见的星”就是今天所说的黑洞。这是我第一次在书中看到有关黑洞的知识。我是一个喜欢读科普书的人，但此后十多年，再也没有从其他科普书和科普文章中看到有关黑洞的介绍。

《每月之星》是我读过的最好的科普书，作者叫陶宏。书中序言的最后写道：“一九四九年一月二十二日，北平停战之日，陶宏写于北大红楼。”

1978年我到北师大天文系读研究生时，在一次谈话中提到《每月之星》这本书和作者陶宏，我的导师之一、时任天文系主任的冯克嘉先生告诉我，陶宏是陶行知先生的儿子。

1996年，我在北师大研究生院工作，有幸与顾明远教授同在一个办公室。中午休息时，我去翻看顾先生书架上的《陶行知文集》。在书中发现许多与《每月之星》相近的内容。

于是我又去翻看《每月之星》这本书的复印件，这才注意到一些我少年时代未曾注意的内容。陶宏在序言中说，这本书采用的是他在给父亲做“小助教”时用过的资料。陶行知先生过世后，他觉得有必要把父亲积累的素材编写成书，并终于编纂出版了《每月之星》这本科普读物。

我当时感到十分震惊，原以为陶行知先生只是一个文科知识和社会知识丰富的教育家，没想到他除去“中晓人和”之外，居然还“上知天文，下知地理”。我看到了一位真正伟大的教育家的形象，他为当代的教育工作者树立了光辉的榜样。细想起来，孔夫子不也是一位百科全书式的人物吗？

我上大学期间，虽然在大三结束的暑假（1965年）自学过广义相对论，但所看的那三本书中均未谈到黑洞。我再次注意到黑洞是在大学毕业之后，分配到哈尔滨东北石油化学所工作期间。1974年的一天，我在所资料室中翻阅杂志，注意到当时的《科学通报》上有王允然先生写的一篇介绍黑洞的高级科普文章，我看得津津有味。

我在中国科技大学上学期间，王先生是物理系的老师，但未教过我，不认识他，只是听说过他的名字。他和科大的一些老师和同学，通过自学攻读钻进了广义相对论的研究领域，并在宇宙学和黑洞领域开展了研究。他们组的成员有的是我上学时的年轻老师，有的是我的同学，这些熟识的名字给我巨大的鼓舞，也进一步加深了我对广义相对论和黑洞的兴趣。

真正开始钻研黑洞，是在我进入北京师范大学，做了刘辽先生的研究生之后。刘辽先生是一位正直的知识分子。抗战期间，他曾就读过空军幼年学校，准备为保卫祖国而战。抗战结束后，他又参加过共产党的外围组织，散发过《挺进报》。解放后，青年刘辽从北大毕业，打算走

科学报国的道路，但却不幸被打成右派。在那些艰难的岁月里，他没有选择放弃，而是一心苦钻理论物理，让自己的思想在爱因斯坦与相对论的天空中翱翔。

当改革开放的春天降临之际，已经掌握了相对论基本理论的刘辽，终于获得了施展抱负的机会，他开始在北师大建立研究相对论的基地，并在国内各地到处宣讲、传播爱因斯坦的理论。我有幸在这个时候考取了刘先生的研究生，在他的带领下逐渐深入到广义相对论和黑洞研究的前沿，从一个黑洞的旁观者，变成了黑洞的探索者。

当我于三十多年前初次跨入黑洞领域的时候，只有少数研究广义相对论的人对黑洞感兴趣，在那里仔细计算、反复推敲、热烈争论。天文学家一般都对此漠然视之，对宇宙中是否存在黑洞不置可否。

三十多年后的今天，情况倒过来了。天文学家大都对黑洞的存在感到乐观，认为宇宙中存在着黑洞，而且不少人认为可能存在大量的黑洞，不断有人发表论文，说这里有黑洞，那里是黑洞，似乎宇宙中处处都有黑洞。

与此相反，原来研究黑洞的理论物理学家，却有许多人感到情况不太妙，在那里争论、反思。包括霍金在内的一些专家认为，原先可能把黑洞想象得太理想化了，真实的黑洞即使存在，也与理论计算出的可能有很大的差别，甚至也可能根本就不存在黑洞。

天文学家对黑洞的乐观看法，产生自天文观测。我们知道，根据计算，一颗太阳质量的恒星，半径70万千米，密度约为每立方厘米1.4克，形成白矮星后，半径缩到1万千米，密度达到每立方厘米1吨左右。如果形成中子星，其半径将缩到10千米，密度达到每立方厘米1亿~10

亿吨。如果形成黑洞，其半径将只有3千米，“平均密度”达每立方厘米100亿吨。

现在，白矮星早已在天文观测中大量发现，约占恒星总数的十分之一。原来认为不可思议的中子星也已发现很多。而从天文学角度来看，黑洞与中子星的半径和密度其实相差不大。因此，黑洞的存在似乎是顺理成章的事了。

另外，类星体巨大能量的来源，星系中心看不见的超大质量，以及引力透镜等，这些观测结论都能用黑洞给出较为合理的解释。

而物理学家对黑洞的反思则来源于黑洞理论造成的信息疑难、奇点疑难等带有根本性的理论困难，以及以往对黑洞的过于理想化的认识。

黑洞究竟是一种什么样的天体，它到底存在不存在？本书将从天文学、物理学和数学的不同角度，来阐释黑洞的来龙去脉，它的几何结构和物理结构，它可能有哪些有趣的效应。本书还将介绍，对黑洞的研究，给物理学带来了哪些收获和重要启示。

黑洞这种从广义相对论和微分几何推演出来的天体，在没有引入任何统计假设的条件下，居然自动导出了温度和熵。也就是说，当时空弯曲到一定程度的时候，物体将自然出现热效应。这种不可思议的结果显示，万有引力与热之间，存在着目前尚不清楚的深刻的本质联系。

对黑洞和弯曲时空的深入思考，进一步把我们的注意力引向了“时间之河”。于是，我们在广义相对论和黑洞理论的基础上，深入探讨了热力学定律与时间性质之间的关系。

我们的“大胆猜测”和“小心求证”也许会勾起不少读者的兴趣，给他

们带来愉悦的思考，并把他们引向持续千年的难题：时间究竟是什么？

作者在写作本书的过程中，谨记一个原则：一本好书，不只要让人相信其内容，更重要的是让人有兴趣去进一步思考、探索 and 发现。

具有高中以上学历的读者不仅可以从本书中学到许多关于广义相对论、黑洞和时间本性的有趣知识，而且能够在不知不觉中提高自己的怀疑能力、创新能力和探究能力。

本书是在清华大学出版社和邹开颜编辑的大力协助下完成的，还得到石磊、朱红莲等编辑的支持；北京师范大学物理系研究生祁景钊、梁桂荣协助打印了书稿，作者在此深表感谢。

作者的科研、教学和写作工作，长期得到国家自然科学基金（如10373003，10773002等）和教育部“精品视频公开课”经费，以及北京师范大学教学经费的支持，在此一并表示深切谢意。

赵 峥

2014年夏于北京

目 录

[引言](#)

[第一章 初窥黑洞](#)

[第二章 恒星演化，走向黑洞](#)

[第三章 弯曲的时空](#)

[第四章 奇妙的黑洞](#)

[第五章 黑洞附近的物理效应](#)

[第六章 探索黑洞的明星——霍金](#)

[第七章 黑洞的信息佯谬](#)

[第八章 奇点——时间有无开始与终结](#)

[第九章 时间测量的疑难与探索](#)

[第十章 千古难题：时间是什么](#)

[主要参考书目](#)

[返回总目录](#)

第一章 初窥黑洞

加尔各答黑洞

我们上来先讲一段题外话。

地球上曾有一个叫黑洞的地方，在印度的加尔各答。那里有一座城堡，印度作为英国的殖民地时，它曾是英军的据点。城堡中有一间不到24平方米的牢房，只有两个小窗，英军用它来关押喝醉酒的士兵，通常可以关三四个人。他们称这间牢房为“黑洞”。

有一次印度爆发了反抗殖民者的民族起义，起义军围攻这座城堡，展开了一场血战，尸横遍野。4天后，战斗结束，愤怒的胜利者把146个俘虏使劲都塞进了这个“黑洞”，关押了10个小时。当时正值最炎热的6月，第二天早晨打开“黑洞”的门时，其中123人已经死去，只剩下23个人还活着。

20世纪90年代在加尔各答开了一次天体物理研讨会，中国科技大学的卢炬甫教授向主办者提出能否参观一下这个“黑洞”。主办方表示抱歉，说这座城堡已经拆掉了。卢教授很惊讶，说怎么会拆掉，如果在我们中国，一定会保存下来作为爱国主义的教育基地。印度朋友说：“哦，你搞错了，当时不是英国人把印度人关到里面，而是印度人把英国俘虏塞到了里面。”

拉普拉斯与米歇尔的“暗星”

今天我们所说的黑洞是相对论预言的天体，然而在19世纪末，欧洲的两两位学者就曾依据牛顿力学预言过黑洞的存在。不过他们当时没有称其为黑洞，只是讨论过这种质量巨大、发光可能很强，但远方观测者又看不见的暗星。

其中一位学者拉普拉斯（1799年）曾在他的巨著《天体力学》和科普书《宇宙体系论》中谈到过这种暗星，他在书中写道：

“天空中存在着黑暗的天体，像恒星那样大，或许像恒星那样多。一个具有与地球同样密度，而直径为太阳250倍的明亮星体，它发射的光将被它自身的引力拉住，而不能被我们接收。正是由于这个道理，宇宙中最明亮的天体很可能却是看不见的。”

然而最早预言这种暗星的还不是拉普拉斯，而是英国剑桥大学的学监米歇尔。他在1784年的一篇论文中谈到，由于恒星发射的光在万有引力作用下光速会越来越慢，如果恒星足够大，大到一定程度，就会使自身发射的光被自身的引力拉回来，于是远方的人就看不见这颗星了。

现在我们先回顾一下牛顿和他的物理理论，然后再来介绍拉普拉斯和米歇尔关于暗星的预言是如何从牛顿理论得出，又如何被否定的。

苹果落地的故事

大家都知道牛顿与苹果落地的故事。这个故事说，牛顿20多岁在乡下庄园里生活的时候，有一次坐在苹果树下思考问题，这时突然有一个苹果从树上落下来，使牛顿一下子想出了万有引力定律（图1-1）。

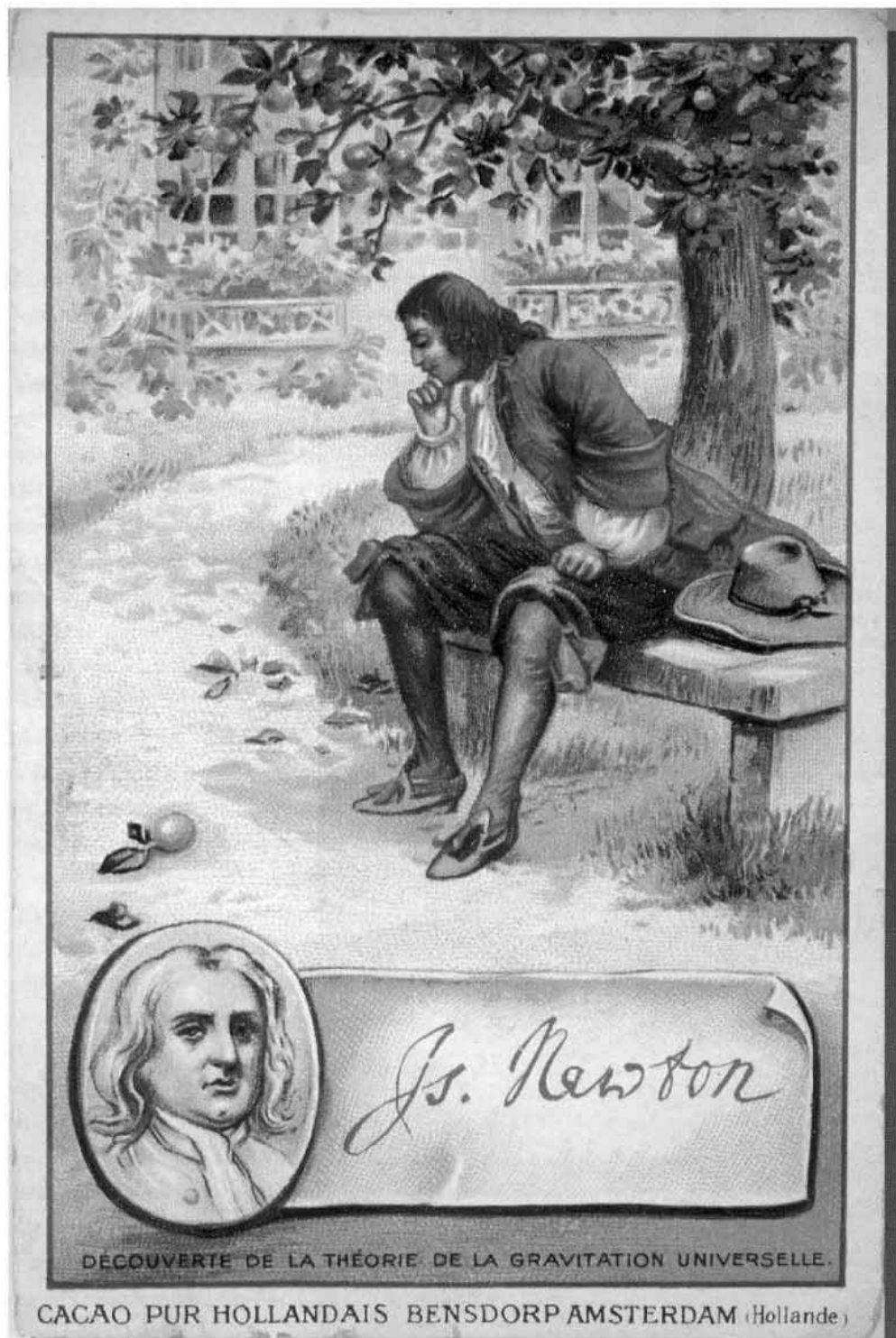


图1-1 苹果落地的故事

这个故事非常有趣，只可惜其真实性很值得怀疑。牛顿生前从来没

有人听到过这个故事，牛顿一死这个故事就冒出来了。而把这个故事传播到全世界的是法国大文豪伏尔泰。

牛顿去世时，恰好伏尔泰被法国当局驱逐出境，流亡在英国。他看到有好几万人给牛顿送葬，深受感动，觉得这个人实在太伟大了，于是去拜访了牛顿的亲属。

牛顿一辈子没有结婚，先是他的妹妹照顾他的生活，妹妹去世后就由他的外甥女照顾他的生活。伏尔泰拜访的就是牛顿的外甥女婿，这位绅士给伏尔泰讲述了苹果落地的故事。才华横溢的伏尔泰把这个故事加以润色演绎写进自己的文章，使它传遍全球。

然而，在此之前外界从未有人听说过这个故事。伏尔泰把它传播开后，有人对其真实性感到怀疑，这时又有几个在牛顿身边生活过的人来说，他们也听牛顿讲过这个故事。

这么重要的一个使牛顿产生伟大灵感的故事，牛顿在世时居然没有和一位学者谈到过。何况中年以后的牛顿并不是一个谦虚的人，他后半生一直在跟别人争夺各项科研成就的发现权，其中很重要的一部分就是和胡克争夺万有引力定律的发现权。牛顿与外人进行了如此多的争吵，怎么一次也没有在争吵中提及这一故事呢？所以，许多人认为，这个故事很可能是牛顿的亲属为了确保他对万有引力定律的发现权而编出来的。

牛顿关于万有引力定律的论述最早出现在他的巨著《自然哲学之数学原理》中，此书是在他45岁时出版的。而苹果落地的故事发生在牛顿23~25岁之间，在乡下庄园里躲避鼠疫的时候。这个故事的出现，把牛顿发现万有引力定律的时间整整提前了20年，那是胡克等人所望尘莫及

的。

童年的牛顿

牛顿的童年是不幸的，他是一个遗腹子，还没有出世，父亲就去世了。他的父亲是一个农民，文盲，而且脾气暴躁，不过据说智力还可以，家里也不算太穷。牛顿的父系家族基本没有文化，母系家族有点文化，但也不高。

后来母亲改嫁给一位60多岁的牧师。牛顿的继父有文化，也有些钱财。但这位“思想道德工作者”本人的道德水平并不高，他不能善待前夫的儿子。

牛顿小时候饱受歧视，与同父异母的弟弟关系很差，只与妹妹关系较好。幼年的牛顿由外祖父母抚养，很少得到母爱。这样的生活状况对牛顿性格的形成造成了极大的影响，致使晚年的牛顿显得尖刻而贪婪。

在这样的家庭条件下，童年的牛顿身体很差，学习也很差，和同学相比他什么都不行，对自己也缺乏信心。此时一个突发事件改变了牛顿的人生。

他们班上有一个小霸王，欺负牛顿，踢了他的肚子一脚，疼得牛顿忍无可忍。愤怒之下，小牛顿不顾自己身单力薄，憋足了全身力气，把那个小家伙揍了一顿。没想到居然打赢了，小牛顿马上信心大增，觉得自己还行啊。他想，既然打架能赢，说不定学习也行。打架产生的自信没有把牛顿引向邪路，反而使他走上了刻苦学习的正轨。于是他的功课大有起色，而且越来越好。

不幸的是，牛顿10岁时，继父又去世了，家中缺少劳动力。上中学时，母亲把他召回家，让他干农活。但他农活儿干不好，也不大上心，

他放的羊把别人家的庄稼吃了不少。还是牛顿的舅舅有远见，觉得自己的外甥别看干农活儿不行，也许在学习上是一块好料。于是说服姐姐让牛顿继续读书。

中学毕业时，牛顿的舅舅又积极活动，使他得以进入剑桥大学学习，并为他从学校争取了补助金。其实此时牛顿的母亲已从前夫那里继承了遗产，并不贫困，但她不愿为儿子掏学费。牛顿能拿到补助金，正合他母亲的心意。

然而，舅舅为牛顿争取到的这点补助金不是白给的，穷学生们必须为富有的同学打扫宿舍，买东西，甚至倒尿壶。这样的大学生涯，对于牛顿后来的性格形成肯定也有负面影响。好在牛顿的求知欲望强烈，而且确实才华出众，他终于以优异的成绩在剑桥毕业，并有幸留校当了教师。

牛顿的丰收年

刚刚留校，英国就闹鼠疫，牛顿不得不到乡下母亲的庄园中去躲避。他23~25岁之间，在那里度过了一年半岁月。不过他没有虚度光阴，这时的牛顿已步入了科研的正轨，他在那里思考、研究了许多问题。据后来他自己说，他的力学三定律、万有引力定律、光的微粒说、色散理论、微积分，都是在那段时间中研究出来的。不过，他没有提到苹果落地的故事。这个故事如果有，就应该发生在这段时间。

牛顿在庄园中度过的这一年半时间，后来被称为牛顿的“丰收年”。

青年牛顿是幸运的，他留校后碰到了伯乐式的导师巴罗。实际上，牛顿的很多哲学和物理思想来源于巴罗。在巴罗的著作中我们可以看到他对绝对空间、绝对时间以及物体运动等许多重要概念的论述。在牛顿的巨著《自然哲学之数学原理》中不难看到巴罗的影子。

当时一位叫卢卡斯的富翁为了支持自然科学研究，给剑桥大学捐了一笔钱，设立了一个叫做“卢卡斯数学讲座”的教席，为数学和物理专业的教授支付薪金。

巴罗是第一任卢卡斯讲座教授，但他看到牛顿的杰出才华后，立刻让贤，把这一讲座教授的位置让给了牛顿。牛顿当时才27岁。牛顿确实身手不凡，他30岁时又当上了皇家学会会员（相当于院士）。

牛顿在《自然哲学之数学原理》（图1-2）中，论述了自己对力学的研究成果，谈到了绝对空间与绝对时间，还谈到相对空间与相对时间。他认为：

“绝对空间，就其本性而言，与任何外部事物无关，它总是相同的和不可动的。相对空间是绝对空间的某个可动的部分或量度.....”

“绝对的、真实的和数学的时间自身在流逝着，而且因其本性均匀地、与任何外部事物并不相关地流逝着，它又可以叫做延续性。相对的、表观的和普通的时间是延续性的一种可感知的、外部的（无论是准确的或不均匀的）借助运动来进行的量度，我们通常就用它来代替真实时间；例如一小时、一个月、一年。”

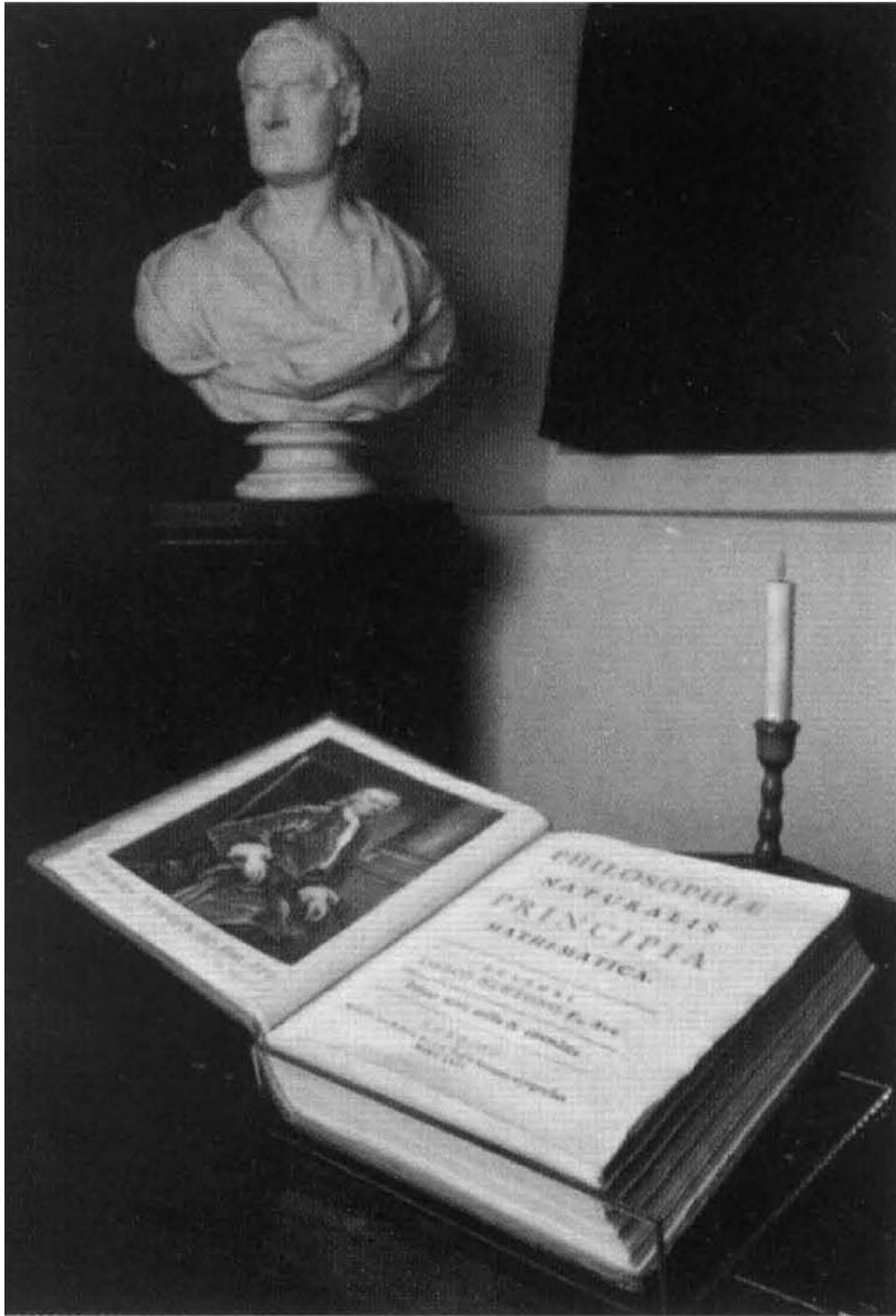


图1-2 《自然哲学之数学原理》第一版

牛顿列出了力学三定律，给出了万有引力定律。这部书奠定了经典力学的基础，其内容之完善，逻辑之严谨都堪称科学著作的楷模，值得

所有热爱物理学的青年学子浏览。

毫无疑问，牛顿是迄今为止最伟大的物理学家之一。

牛顿、胡克与万有引力定律的发现

牛顿时代的天文学家和物理学家，都熟知开普勒通过精细的天文观测得到的行星运动三定律（图1-3）：

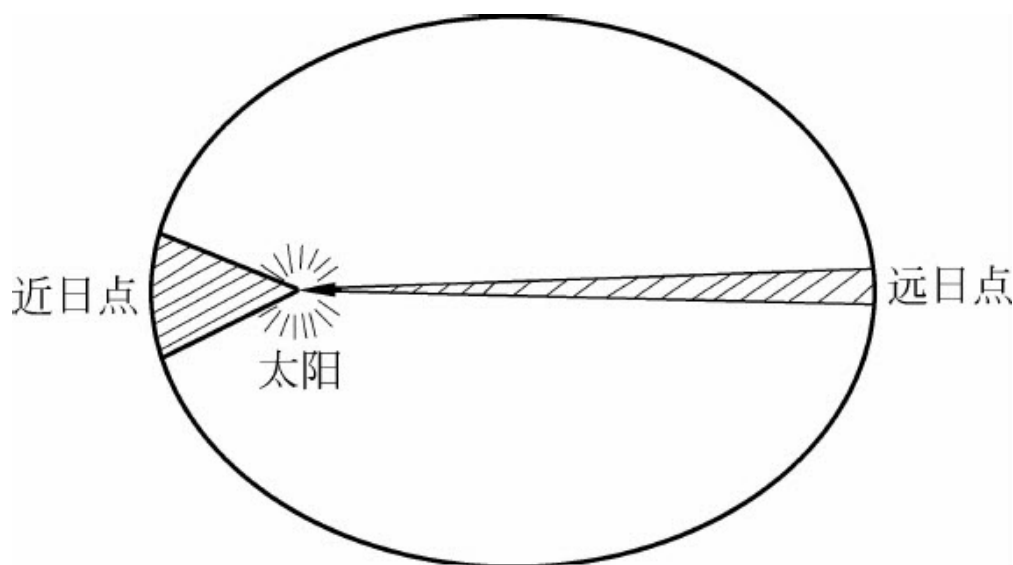


图1-3 开普勒的行星运动定律

第一定律：行星绕日运动的轨道是一个椭圆，太阳位于椭圆的一个焦点上。

第二定律：行星的矢径在单位时间内扫过的面积相等。

第三定律：行星绕日运动周期的平方，与椭圆轨道半长轴的立方成正比。

牛顿和胡克等人几乎同时认识到，如果开普勒第三定律成立，则太阳和行星间的引力就一定与距离的平方成反比。问题是，与距离平方成反比的力是否一定能导致行星运动轨道是一个椭圆？

有一次，哈雷与胡克等人在一起议论这个问题，胡克说，与距离平方成反比的引力，一定能导致行星运动轨道是椭圆，而且声称他证明过这一点。哈雷想看看他的证明，胡克不肯。

于是哈雷又去请教牛顿，牛顿也说，与距离平方成反比的引力一定能导致椭圆轨道，自己也证明过。哈雷说他想看一下牛顿的证明，牛顿当场找了半天，也没有翻出来。不过牛顿答应再证一遍给哈雷看。不久之后，哈雷收到了牛顿的来信，看到了牛顿的证明。

当牛顿发表《自然哲学之数学原理》一书时，遇到了麻烦。胡克说你书中的万有引力定律是我先发现的，于是二人争吵起来，导致书无法出版。牛顿不得不做出让步，在书的序言中被迫写上胡克也是“万有引力与距离平方成反比”的发现者之一。不过，给出万有引力定律的正确而完整表达式（1.1）的人，肯定是牛顿。

$$F = G \frac{Mm}{r^2} \quad (1.1)$$

牛顿与夫莱姆斯梯德的冲突

除去与胡克争夺万有引力定律的发现权之外，牛顿还与天文学家夫莱姆斯梯德发生了争吵。牛顿在《原理》一书中引用了夫莱姆斯梯德的许多天文观测数据。《原理》一书再版时，牛顿听说夫莱姆斯梯德又有了许多新数据，就向他要，夫莱姆斯梯德不给。牛顿当时已经坐上了皇家学会会长的宝座，傲气十足，就命令夫莱姆斯梯德把数据交出来。夫莱姆斯梯德就是不给，牛顿也没有办法。于是牛顿又想了一招，让他的好友哈雷去向夫莱姆斯梯德借这批数据，哈雷拿到数据后，按照牛顿的主意把这批数据用在自己写的一本书上。牛顿想，哈雷这本书一出版，自己就可以光明正大地引用这批数据了。不料此事被夫莱姆斯梯德觉察到了，他就把哈雷告上了法庭，法庭裁决不许哈雷出版这本“剽窃的著作”。

牛顿一气之下，删掉了《原理》一书中原已引用过的来自夫莱姆斯梯德的所有数据。

夫莱姆斯梯德非常愤怒，此后不再和牛顿讲话。他认为“牛顿是一个阴险的，野心勃勃的、极其贪婪的沽名钓誉者，根本不能容纳不同意见……”，不过他最后还是表现得比牛顿有度量，较为公正地评价了牛顿，“……归根到底是个好人，但生性多疑。”

牛顿与莱布尼茨的争吵

最著名的冲突发生在牛顿与德国数学家莱布尼茨之间，二人争夺微积分的发现权。现在已经清楚，二人是各自独立创建微积分的。牛顿是在研究速度时发现的，莱布尼茨则是在研究切线斜率时发现的。牛顿的发现时间要早一点，但他公布发现的时间晚于莱布尼茨。

当时二人公开争吵，双方的朋友都出来帮忙。牛顿的朋友写文章说是牛顿先创建的，莱布尼茨的朋友则发表文章说是莱布尼茨先创建的。不过后来发现，不少牛顿朋友所写文章实际上是牛顿本人先写好，然后以他朋友的名义发表的。

在争吵中莱布尼茨犯了一个错误。当时英国的科学比德国发达，于是莱布尼茨请求英国皇家学会出面成立一个“公正的委员会”来裁决一下，究竟是谁先发现了微积分。

让莱布尼茨没有想到的是，作为皇家学会会长的牛顿就敢胡来，牛顿以会长身份指定了一个全部由自己的朋友组成的“公正的委员会”，这一委员会“裁决”莱布尼茨剽窃，牛顿还私下为这个委员会起草了裁决书。

委员会裁决之后，牛顿还觉得不解气，又化名写了一篇文章，回顾这件事情的来龙去脉，以彻底把莱布尼茨搞臭。据说，莱布尼茨非常伤心，不久就去世了，牛顿则沾沾自喜。

莱布尼茨是一位才华横溢的学者，他不仅研究数学，还研究物理、哲学、历史、法律、神学和外交。

莱布尼茨与牛顿不仅在微积分的发现权上产生冲突，而且还在哲学和时空观上进行过争吵。牛顿说存在一个“绝对空间”和一个“绝对时间”，莱布尼茨说根本就没有“绝对空间”，也没有“绝对时间”，一切都是相对的。“空间”不过是物体相对位置和方向的表现。时间不过是事件发生的先后顺序的排列。没有物质，就没有时间和空间。



绘画: 张京

神坛上下的牛顿

牛顿的科学发现基本上都是在青年和中年时期做出的，他为此付出了艰苦卓绝的劳动。他的秘书曾经回忆牛顿在写作《原理》那段时期的情况：“他从不作任何娱乐和消遣，他不骑马外出换空气，不散步，不玩球，也不做任何其他运动。认为不花在研究上的时间都是损失。他常常工作到半夜三更，往往忘记吃饭，当他偶尔在学院的餐厅出现时，常常穿一双磨掉了后跟的鞋，袜子乱糟糟，披着衣服，头发也几乎不梳。”

牛顿的伟大成就对物理学产生了深远的影响。然而中年以后的牛顿，不是一个招人喜欢的人，他很不谦虚，不停地与人争吵，在自然科学上不再有建树。

有人说晚年的牛顿特别想发财，最后终于当上了造币厂的厂长。不过，牛顿确实了不起，他在管理金融时提出的“金本位”主张，对经济产生了深远影响。英镑和美元最硬的时候，都是在它们的币值与黄金挂钩的时候，脱钩后，就不可避免地走向了衰落。所以，“金本位”的主张，值得崛起中的大国深思。

牛顿有一句名言：“我之所以有成就，是因为站在了巨人的肩膀上。”许多人以为这是牛顿谦虚的表现，事实并非如此。这句话是在牛顿给胡克的信中出现的。胡克是个矮子，而且有点驼背。牛顿是想说，我能够做出成就是因为我站在了笛卡儿等巨人的肩膀上，跟你胡克这个矮子无关。

晚年的牛顿也还是有谦虚的时候，不过他敬畏的对象不是某个人而

是自然界。他曾说过：“我不过像一个在海边玩耍的小孩，时而发现一块光滑的石子，时而发现一个美丽的贝壳，但真理的广阔海洋，却还在我的面前有待发现。”

如何得出暗星的预言

拉普拉斯和米歇尔是依据牛顿的万有引力定律和力学第二定律预言这类暗星的。他们算出暗星形成的条件是：

$$r \leq \frac{2GM}{c^2} \quad (1.2)$$

其中 r 、 M 分别是恒星的半径和质量， G 和 c 分别是万有引力常数和真空中的光速。不过他们当时还不知道 c 是一个常数，更不知道光速是自然界最高的极限速度。他们以为光速和一般质点的速度一样，在外力下会依据牛顿第二定律而变化。他们认为，如果式（1.2）的条件被满足，这颗恒星的光就会被自身的引力拉回去，成为外界看不见的暗星。

我们从经典力学的动能与势能关系很容易推出此式。设光子的质量为 m ，则它的动能为 $\frac{1}{2}mc^2$ ；光子位于恒星表面时，万有引力势能为 $\left(-G\frac{Mm}{r}\right)$ 。光子抵达远方的条件是它的动能能够克服势能，如果动能小于势能，远方的观测者就接收不到这个光子了。所以形成暗星的条件可以从

$$G\frac{Mm}{r} \geq \frac{1}{2}mc^2 \quad (1.3)$$

导出，于是我们得到了式（1.2）。

从今天看来，上面的论证有两方面错误，第一是按照狭义相对论，光子动能不是 $\frac{1}{2}mc^2$ ，而是 mc^2 ，而且真空中的光速 c 是一个常数，不会在外力作用下改变。第二是万有引力与一般力不同，是一种几何效应，万有引力定律只不过是爱因斯坦广义相对论的一个近似。有趣的是，这

两方面的错误相互抵消，式（1.2）与后来从相对论得出的黑洞形成条件恰好一致。

$$r_g = \frac{2GM}{c^2} \quad (1.4)$$

式中， r_g 与 M 分别为黑洞的半径和质量， r_g 又称星体的引力半径。

高傲的拉普拉斯

拉普拉斯是当时世界上最杰出的数学家和天体物理学家，他的五卷巨著《天体力学》受到拿破仑的关注。这位关心科学技术的独裁者在翻阅了拉普拉斯的《天体力学》后，召见了自己的这位同胞，问他：“这部书中怎么没有提到上帝的作用？”拉普拉斯高傲地回答：“我不需要这个假设。”

拉普拉斯在《天体力学》第一版（1796年）和第二版（1799年）中都谈到了自己预言的暗星，但在1808年出版的第三版中却悄悄地删除了有关暗星的叙述。

这是因为在第二版和第三版的出版之间，托马斯·杨完成了光的双缝干涉实验，这表明光是波而不是微粒。拉普拉斯感到自己建立在牛顿微粒说基础上的暗星预言，看来不对了。

托马斯·杨的功绩

英国的托马斯·杨小时候是个神童，长大后成了才子。托马斯·杨2岁时就能读书，4岁的时候把《圣经》通读了两遍，到14岁的时候就学会了拉丁语、希腊语、法语、希伯来语、意大利语、阿拉伯语、波斯语等，会多国的语言。他先是学医，研究近视眼，弄清了散光的原因；又对光学感兴趣了，完成了双缝干涉实验，证明了光是波动，而且是横波，还提出了颜色的三色理论。

托马斯·杨在十来个领域都有贡献。特别滑稽的是，他对考古学也有贡献，他把古埃及的罗塞塔石碑上的文字破译了几个，古埃及文研究的第一次突破就是他首先认出了几个字，当然没有全部突破，但也是一个很重要的进展。

光的波动说和微粒说经历过长时期的争论。最先是笛卡儿、胡克、惠更斯等人提出波动说，认为光是波动。这些人都比牛顿资格老。牛顿在剑桥大学任教后，提出光的微粒说，认为光是微粒。他把论文投给英国皇家学会的会刊，遭到学会干事长胡克的否定。胡克认为光是波动已有定论，牛顿的论文是胡扯，他把牛顿的论文退了回去。

牛顿一气之下，从此以后不再给皇家学会会刊投稿，所以牛顿一生论文发表得很少，他的主要成就都刊登在45岁时出版的《自然哲学之数学原理》和65岁时出版的《光学》这两本巨著中。世间后来流传的一些牛顿的论文如“论运动”等，都截自牛顿与别人的私人通信。

在《光学》一书中，牛顿叙述了自己对光学的研究成果，阐述了光的微粒说。由于他在力学方面的巨大成功，大家相信他对光的论述也应

该是正确的，因为他是伟人。

何况惠更斯学派一直未能解释清楚，作为波动的光为什么一直观察不到干涉现象。这样，牛顿的微粒说就压倒了惠更斯的波动说，这一压倒就是一百多年，直到托马斯·杨完成光的双缝干涉实验。

由于牛顿是伟人，是英国民族的骄傲，所以托马斯·杨在论述波动说时小心翼翼。他一再肯定牛顿的伟大，一再申明自己仰慕牛顿的大名，只是希望自己能对牛顿的失误作一点修正，使牛顿的物理理论更加完美。

但这也不行，托马斯·杨受到来自学术界和社会的双重压力，有人说他的文章“没有任何价值”，他的双缝干涉实验根本“称不上是实验”。托马斯·杨的论文无法发表，只好自费印了一些小册子，但也只卖出了一本。不过，真理是否定不了的。牛顿的微粒说无法解释双缝干涉实验，光的波动说最后还是战胜了微粒说。

奥本海默的暗星

光的波动说战胜微粒说之后，米歇尔和拉普拉斯建立在微粒说基础上的暗星预言，逐渐被人们淡忘。

学术界再次谈论暗星是100多年以后的事。1939年美国物理学家奥本海默和施耐德在研究中子星的时候，用爱因斯坦的广义相对论再次论证了存在暗星的可能性。

广义相对论可以看作万有引力定律的发展和推广。这一理论认为，万有引力不是一般的力，而是时空弯曲的表现。

奥本海默等人用广义相对论算出的暗星形成条件如式（1.4）所示，与米歇尔、拉普拉斯给出的条件一致。但是他们依据的理论已远非拉普拉斯等人依据的经典力学可比，他们认为暗星的存在不是万有引力把光拉了回来，而是星体质量造成的巨大时空弯曲，把光束束缚在了暗星内部，无法逃逸。

然而，暗星的“密度”大得几乎无法让人相信，太阳如果形成暗星，半径会从70万千米缩小到3千米，密度似乎会达到每立方厘米100亿吨。这真是一个让人无法接受的“天文数字”！

当时已知密度最大的物质是白矮星上的物质，其密度也不过每立方厘米1~10吨。更为可怕的是，暗星内部的物质似乎都会缩到中心的一点上，形成密度和时空曲率都为无穷大的“奇点”。包括爱因斯坦在内的绝大多数物理学家都不相信宇宙间真会有这样的暗星存在。

不久之后，奥本海默受命主持原子弹的研制，对暗星的研究再次中

断。

1964年前后，美国相对论专家惠勒重新研究了奥本海默的暗星形成理论，并用美国核试验基地的大型计算机作了恒星在万有引力作用下塌缩的模拟计算，确认了中子星塌缩真的会形成暗星。他把这一喜讯告诉了奥本海默，然而当时奥本海默由于被人诬陷泄露原子弹机密，而遭到联邦调查局的反复审查，这使他情绪低落，失去了继续研究暗星的兴趣。

不过惠勒的工作终于引起了相对论界的重视，对暗星的探索重新启动，惠勒还给这种暗星起了个专用的名字叫“黑洞”，于是黑洞一词逐渐传播开来。

第二章 恒星演化，走向黑洞

赫罗图

在远离城市的郊外，仰望晴朗的夜空，可以看到万点繁星。如果使用望远镜，就能更好地欣赏那五颜六色的恒星。它们都是遥远的太阳。

天文学家根据恒星的发光本领（光度）和温度，绘制出一种图，是用制作此图的天文学家赫茨普龙和罗素的名字来命名的，称为赫罗图（图2-1）。此图的横坐标表示恒星的温度，纵坐标表示它们的光度。

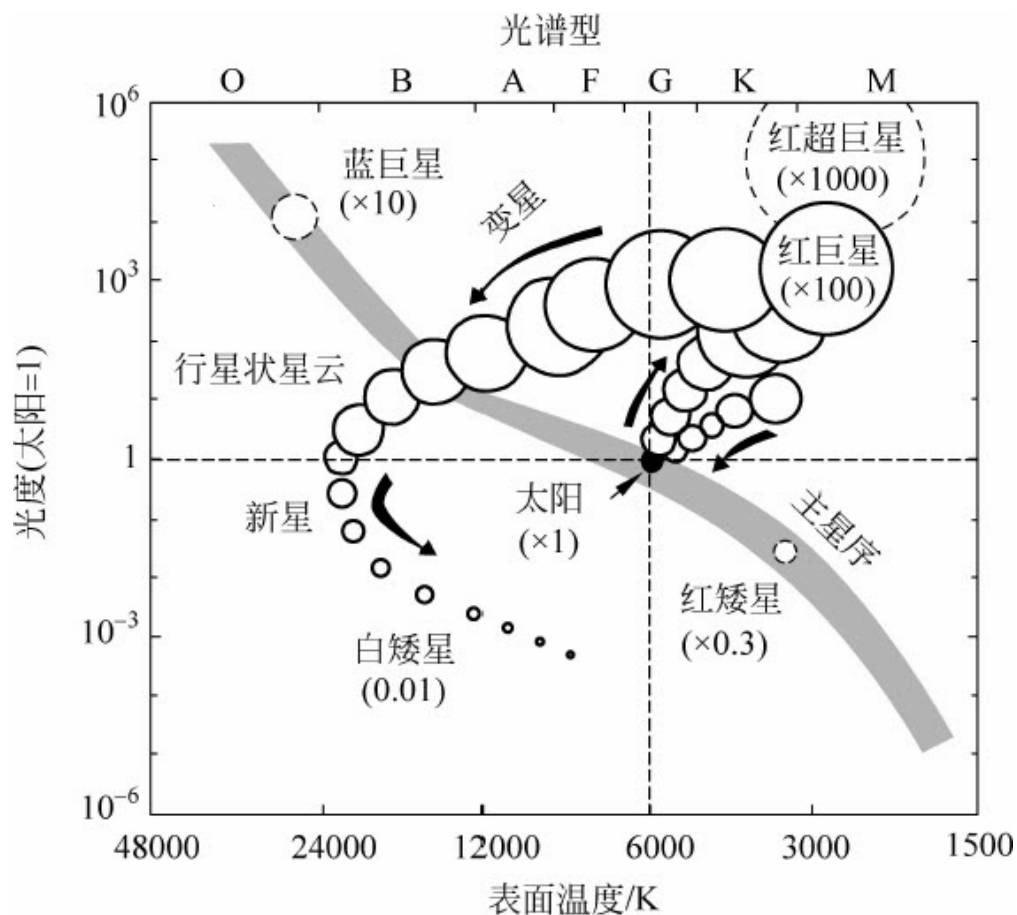


图2-1 赫罗图

恒星的温度可以用它们发射的光的光谱来确定。温度较低的恒星，发射的光波长较长。4000K左右的恒星主要发红光，呈现红色，温度更低的恒星则发射红外线、微波或波长更长的无线电波。6000K左右的恒星主要发黄光，呈现黄色，例如我们的太阳；1万度左右的恒星主要发白光，呈现白色。温度更高的恒星则发射蓝光、紫外线甚至X射线。

肉眼所见的恒星亮度不仅决定于星体自身，而且决定于它们离我们的远近。同样种类和大小的恒星，离我们越近，显得越亮。我们感兴趣的不是这种视亮度，而是恒星的“绝对亮度”。通过天文测量，可以知道恒星离我们的距离。我们把所有恒星都折算到离我们10个秒差距的距离（1秒差距 \approx 3.259光年），它们这时候显示给我们的亮度，称为绝对亮度，它反映恒星的真实发光本领——光度。

赫茨普龙和罗素根据不同恒星的光度和颜色，把它们一一标记到赫罗图中。结果显示，大多数恒星都分布在从左上角到右下角的对角线附近，形成一个恒星比较集中的带，称为主星序。主星序上的恒星称为主序星，我们的太阳就是一颗主序星。还有一些恒星分布在主星序的两侧，有体积较大温度较低的红巨星，有体积较小温度较高的白矮星，此外还有蓝巨星、红矮星等。

吻我一下吧，仙女！

根据恒星光谱中某些谱线的特征，天文学家把恒星分成O，B，A，F，G，K，M等若干种光谱型，标记在赫罗图上方。这些光谱型的排列顺序较为难记，有人编了一个笑话来帮助记忆，说有一个青年第一次来到天文台，用望远镜来看夜空，那五颜六色的天体让他惊叹，不禁喊了一声：

Oh, be a fine girl, kiss me!

“哦，真像一个美丽的仙女，吻我一下吧！”这句话每个英文单词的第一个字母恰好表示按顺序排列的光谱型。

研究表明，恒星在赫罗图中的位置，显示它们的不同演化阶段。主星序上的恒星比较年轻。老年的恒星逐渐离开主星序，先演化成红巨星或超红巨星，再进一步演化成白矮星或者其他致密星。

爱丁顿的贡献

最初，物理学家们认为恒星发光发热的能量完全来自引力势能。他们认为在气体星云收缩为恒星时，气团的引力势能会转化为热能，使恒星温度升高，发光发热。不过，他们认为，此后维持这一发光发热过程的能量依然来源于引力势能，来源于恒星物质的继续收缩。也就是说，引力能是恒星热能和光能的唯一来源。著名物理学家开尔文和亥姆霍兹等人就持这种观点。

然而，后来发现恒星的寿命很长，达到几十亿年，恒星物质的引力势能远不能维持如此长时间的发光发热。于是英国天体物理学家爱丁顿提出，恒星的能量源泉是核的聚变反应（即通常所说的热核反应），是4个氢核聚合成氦核的聚变反应。

当时核物理学尚不发达，许多核物理学家认为氢核（质子）所带的正电荷会同性相斥，它们不可能聚合到一起。要使质子靠近，需要给它们提供足够的动能，也就是说，恒星温度要非常高，而当时估计的恒星温度远没有这样高。

针对核物理学家认为恒星温度不够高，不可能形成氢聚合成氦的热核反应的观点，爱丁顿高傲地回答：“我们不同那些说恒星温度不够高的批评者争辩，我们只告诉他们，往前走，去找到为什么会有更高温度的理由。”

历史表明，爱丁顿是对的，恒星的收缩的确可以使其中心的温度非常高，同时压强非常大，使得质子的动能达到足以克服它们之间的静电斥力，相互靠近而发生热核反应的程度。而且，随着核物理学的发展，

人们认识到，当质子、中子等核子趋近到 10^{-15}m （原子核大小）的时候，会出现一种远比静电斥力强大得多的吸引力——核力（即强相互作用力）。正是这种核力把质子、中子聚拢在一起，形成稳定的原子核。

爱丁顿使学术界认识到，恒星的能源不是引力势能，而是聚变反应释放的核能。万有引力势能的作用仅限于“点火”，即原来温度不高，不会产生聚变反应的恒星物质，在万有引力作用下收缩，引力势能转化为热能，使恒星温度不断升高，压强不断增大，直到恒星中心部分的温度和压强达到了诱发热核反应的程度，完成聚变反应的点火。此后，恒星发光发热的能源就不再是引力能，而是核能了。

现在知道，以太阳为代表的主序星的热核反应为



即4个氢核（质子），聚合生成一个氦核（由两个质子和两个中子组成）、两个正电子 e^+ （与通常我们熟悉的电子类似，只不过带的是正电荷）和两个中微子 ν ，并释放出核能。

霍伊尔的贡献

天体物理学家霍伊尔发展了爱丁顿的思想，他首先认识到氦元素还能发生进一步的聚变反应，生成碳元素、氧元素；碳和氧等又可再进一步聚变，生成更重的元素，例如铁、硅等。这不仅解释了恒星发展各个阶段的产能机制，例如白矮星、中子星的形成，以及超新星爆发过程，而且解释了宇宙中重元素的来源。

大家知道，宇宙初期只存在氢和氦两种元素，但今天的宇宙中存在各种重元素（铁、硅等），这些重元素从哪里来？天体物理学家原本不清楚，霍伊尔的工作解开了这一秘密。不过最初的研究是令人沮丧的。

人们发现，两个氦核聚合的生成物（4个质子与4个中子），或者一个氦核与一个氢核聚合的生成物（3个质子与2个中子）均不稳定，这样的聚变反应不可能发生。3个氦核聚合在一起生成的碳（6个质子与6个中子）倒是稳定的，但3个核同时碰在一起的概率很低，这样的反应似乎更难发生。这一严重困难被霍伊尔解决了。

霍伊尔猜测，碳核可能存在一种激发态，其能量恰好与3个氦核加起来的总能量相等，这时在3个氦核与激发态碳核之间会发生一种“共振反应”，使聚合概率大大提高。生成的激发态碳核又会很快跃迁到基态，形成稳定的碳核。这样氦聚合成碳元素的聚变反应就得以进行了。

一些核物理学家最初不相信霍伊尔的猜测，但他们查找后，真的发现存在这种碳的激发态，确认了“共振反应”的存在。大家终于明白了，通过“共振反应”，氦可以进一步聚合生成碳，释放出大量核能。而且碳还可以与氦再进一步聚合成氧。

研究表明，氧还可以再进一步与氦聚合……这样一步步聚合下去，生成各种重元素。各种元素与氦或氢进一步聚合，生成更重的元素，都是可以进行的。各级聚变反应就像一架天梯，每种元素的核与一个氦核或一个氢核聚合，就迈上一个梯级，聚合成更重元素的核。

不过，这架天梯的最下面两级却是残缺的，即两个氦核的聚变不会发生，氦核与氢核的聚变反应也不会发生，天梯中原子量为8或5的梯级都不存在。不过，再往上，天梯就比较完美了，聚变反应可以步步升级，生成的重元素的种类也越来越多。

我们看到，爱丁顿与霍伊尔对天体物理学的贡献是巨大的。不过，伟人也会犯错误，后面我们会看到爱丁顿与霍伊尔的失误。

恒星的演化

现在我们就来简介一下恒星的演化历程，它们如何形成，如何成长，又如何衰老，如何消亡（图2-2）。

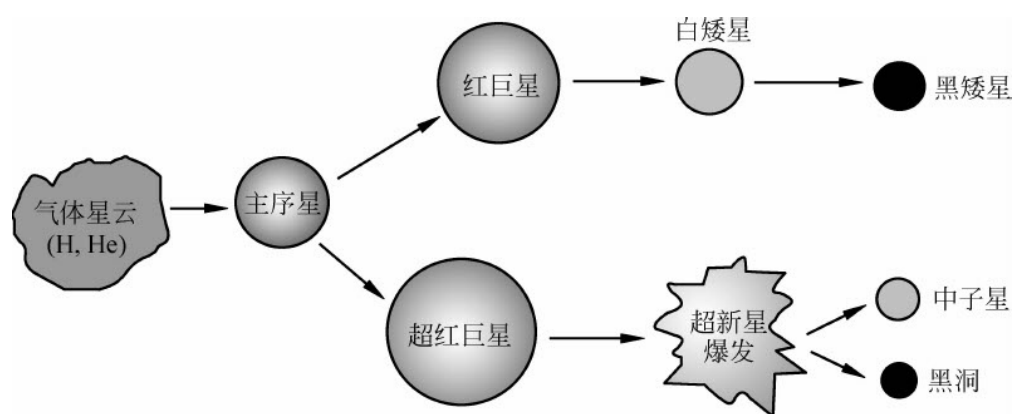


图2-2 恒星的演化

宇宙刚诞生的时候，处在温度极高的状态，最初形成的元素是氢。氢核在高温下发生聚变反应，形成氦核。随着宇宙的不断膨胀，气体的温度逐渐降低，氢聚合成氦的热核反应逐渐停止下来，这时宇宙中的元素大约有70%多的氢和20%多的氦。

这种混合气体不是绝对均匀也不是绝对静止的，随着涨落变化，气体开始聚集成团，并在万有引力的作用下逐渐收缩。在收缩过程中，万有引力的势能逐渐转化成热能，这些气团的温度开始升高，大的气团的中心部分温度可达几千万度，上亿度，压强也可达几千亿个大气压。

在这种高温高压状态下，气团的中心重新开始了氢聚合成氦的热核反应，发出大量的光和热，成为年轻的恒星，于是恒星诞生了。

这些年轻的恒星处在赫罗图的主星序中，称为主序星。我们的太阳

就是一颗主序星。恒星将在主序星阶段持续生存几十亿年到几百亿年。在这个阶段，恒星内部的热核反应相对稳定，发出稳定的光和热。

当主序星内部的氢基本烧完，基本转化成氦以后，外层的氢开始燃烧（聚变），这时恒星开始膨胀，温度也略有降低，成为体积庞大的红巨星或超红巨星。

红巨星和超红巨星的中心部分聚集着大量的氦，这些氦的温度逐渐降低，在万有引力作用下，会进一步收缩，并点燃由氦聚合成碳和氧的聚变反应。

质量小于8个太阳质量的主序星会演化成红巨星，红巨星再演化成由碳和氧两种元素构成的白矮星，它们靠电子之间泡利不相容原理的斥力支撑，成为高密度的白矮星。白矮星密度可达每立方厘米1~10吨。

质量大于8个太阳质量的主序星会演化成超红巨星，超红巨星的中心部分在聚变反应形成碳和氧后，由于万有引力巨大，电子间的泡利斥力支撑不住万有引力的挤压，恒星不会停留在白矮星状态，还会发生进一步的聚变反应，形成铁、硅等重元素，并发生猛烈的爆炸，即超新星爆发，最终形成中子星、黑洞，或全部炸飞而不留残骸。

几种恒星的比较

总之，恒星演化的晚期会经过红巨星（或超红巨星）阶段形成白矮星、中子星或黑洞。

研究表明，剩余残骸质量不超过钱德拉塞卡极限（ $1.4M_{\odot}$ ， M_{\odot} 为太阳质量）的恒星会形成白矮星。

质量超过钱德拉塞卡极限的恒星会形成中子星，这是一种靠中子间泡利斥力支撑的恒星，密度高达每立方厘米1亿~10亿吨。质量超过奥本海默极限（ $2\sim 3M_{\odot}$ ）的恒星，自然界中没有任何力可以支撑，将塌缩成黑洞。

表2-1以太阳为例给出了几种恒星密度与半径的比较。太阳半径70万千米，密度与水差不多，约为每立方厘米1.4克。

表2-1 恒星密度与半径的比较

	太阳	白矮星	中子星	黑洞
半径	70 万千米	1 万千米	10 千米	3 千米
密度	1.4 克/立方厘米	1 吨/立方厘米	3 亿吨/立方厘米	100 亿吨/立方厘米

太阳演化成白矮星后，半径缩小到1万千米，密度约为每立方厘米1吨。如果形成太阳质量的中子星，半径约为10千米，密度为每立方厘米1亿~10亿吨。太阳质量的黑洞，半径为3千米，“密度”约为每立方厘米100亿吨。图2-3给出了各种恒星大小的比较。

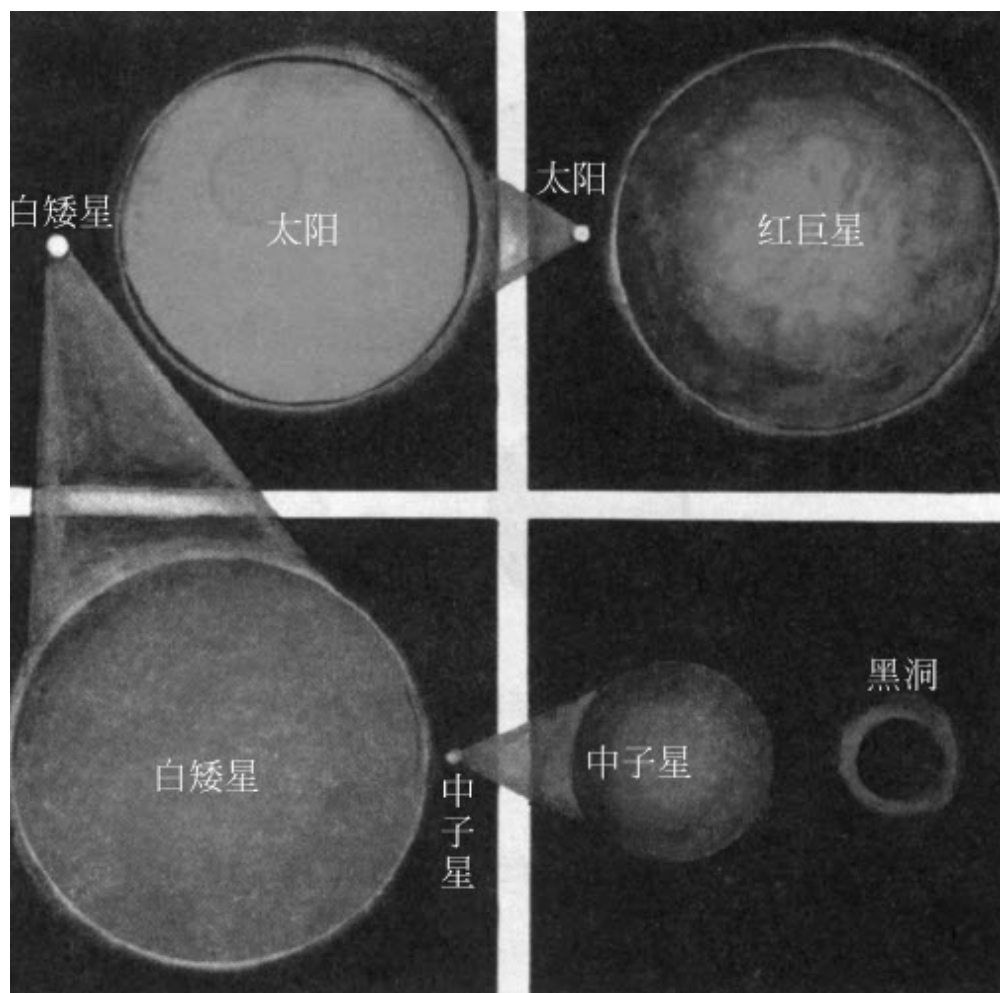


图2-3 恒星尺度比较图

这里有两点需要说明。首先，太阳演化的最终结局是白矮星，不可能是中子星或黑洞。质量最小的中子星或黑洞要由8个以上太阳质量的主序星演化而成。

其次，下面我们会谈到，黑洞内部基本上是真空，物质都聚集在黑洞中心、体积为无穷小的奇点上，所以谈论黑洞的密度实际上没有意义。

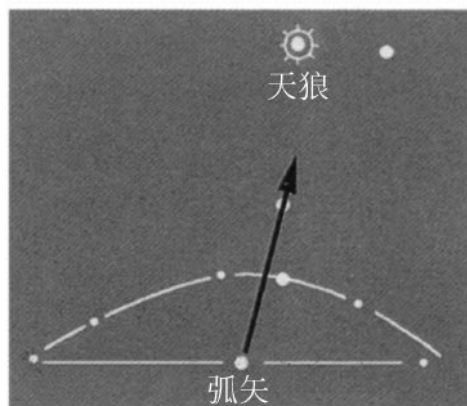
弧矢射天狼

白矮星首先是在天文观测中发现，然后才由理论解释清楚的。

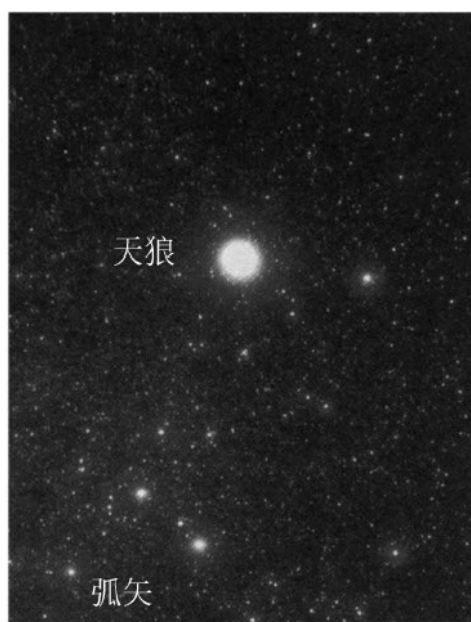
冬季黄昏以后，在天空的东南方有一颗很亮的恒星，那就是人类看到的除太阳外视亮度最大的恒星，中国人称为天狼星。西方人称为大犬座 α 星。

古希腊人按照他们的神话，把天空中的繁星分成几十个星座，例如猎户座、英仙座、仙女座、大熊星座等，每个星座中最亮的一颗称为该星座的 α 星，其次为 β 星，再次为 γ 星等。大犬座 α 星就是大犬座中最亮的一颗恒星。

中国古代认为天狼星代表侵略，天狼星的左下方有一组星，连起来好像弓箭，中国人称为“弧矢”，“弧矢”射“天狼”，表示反击侵略（图2-4）。



(a)



(b)

图2-4 弧矢射天狼

屈原的诗中有“举长矢兮射天狼，操余弧兮反沦降”。苏东坡的诗中也有“会挽雕弓如满月，西北望，射天狼”。这是因为北宋时主要敌人是西夏，西夏在大宋国的西北方。另外，从天空看，虽然天狼星始终出现在天空的南部，但它位于弧矢星的西北。

白矮星的发现

大家知道，天空中的星体，除去太阳、月亮和金、木、水、火、土五颗行星外，它们在天空的相对位置都不变化，所以古人称它们为恒星。也正是因为恒星间的相对位置不变，它们才能形成固定的星座。

实际上，肉眼所见的恒星，都是银河系中的星体，都在一刻不停地围绕银河系的中心转动。不过这些恒星都离我们太远了，最近的也距离我们4光年以上，也就是说光从那里走到地球需要4年以上的时间。由于它们离我们太远，所以我们不可能在较短的时间内察觉到它们的位置变化。想要明显看出它们位置的变化，一般要几万年才行。

图2-5给出了北斗七星现在以及十万年前和十万年后的相对位置。如果今天让一位天文学家冬眠，十万年后再让他苏醒过来，恐怕他那时要辨认出天上群星的名字就非常困难了。

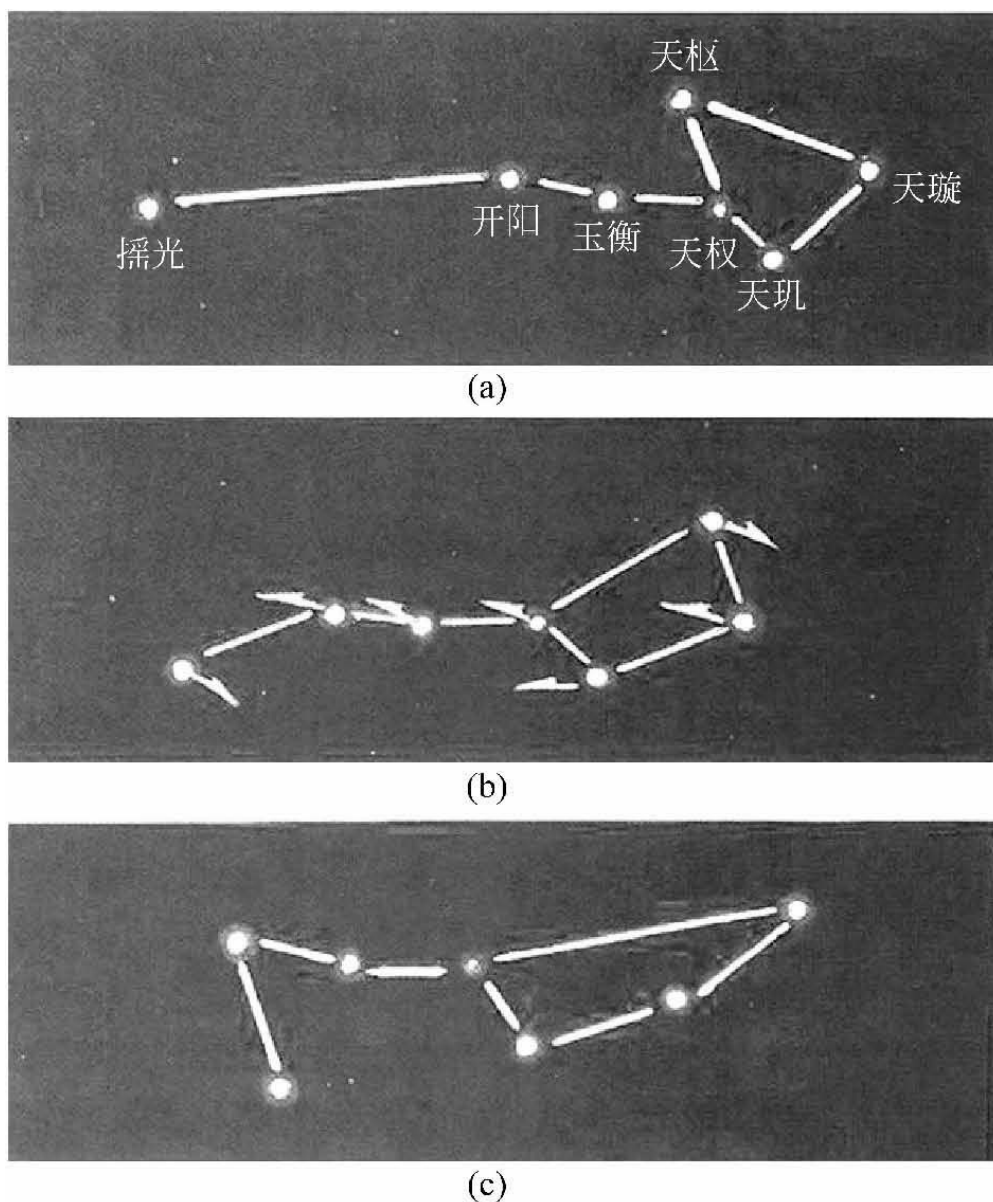


图2-5 北斗七星在天空中位置的变化

不过，天狼星离我们比较近，大约9光年。由于它离我们近，所以它的位置变化就比较容易发现。天文学家在1834年就发现它在绕一个小圆圈作周期运动。它为什么要转圈呢？

当时的物理学家已经知道万有引力定律，那必定是因为天狼星还存在一个看不见的伴星，它们两个在围绕共同的质心转动。有人打了一个

比方说，这就像一对青年男女在跳舞，男孩子穿一身黑礼服，女孩子穿一身白色纱裙。当灯光暗下来的时候，人们就看不见那个男孩子了，只看见一个女孩子在转圈。她为什么转呢？因为有个东西在吸引着她。

28年后，天狼星的这颗伴星被观测到了，称它为**天狼B星**，而天狼星本身则称为**天狼A星**（图2-6）。它们组成一个有两个太阳的太阳系——双星系。实际上，像我们的太阳系这样只有一个太阳的太阳系是少数，大多数太阳系都有两个以上的太阳。有两个太阳的称为双星，有多个太阳的称为聚星。

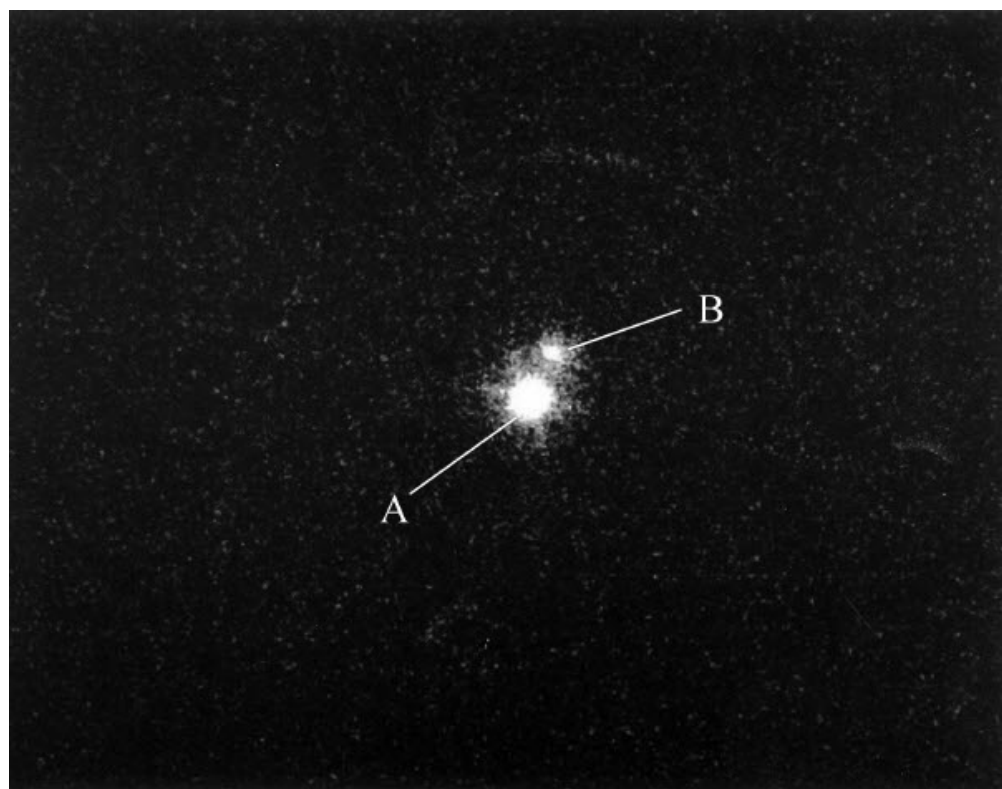


图2-6 天狼A星与B星

天狼**B星**个子很小，发白光，温度约2万度。令人惊讶的是它的密度极高，大约为每立方厘米2.5吨。我们的太阳系中不存在这样的物质，地球上密度最大的物质也不过每立方厘米几十克。天文学家们觉得天狼

B星的密度太惊人，简直不可思议！

经过一段时间的研究，人们逐渐明白了，构成天狼**B**星的是一种特殊物质。

从主序星到红巨星

图2-7是我们观测到的一颗红巨星——心宿二。它是中国古代命名的二十八宿中“心宿”里的一颗，质量比太阳略大一点。从图2-7中可以看出，它非常巨大，能把火星轨道都包进去。

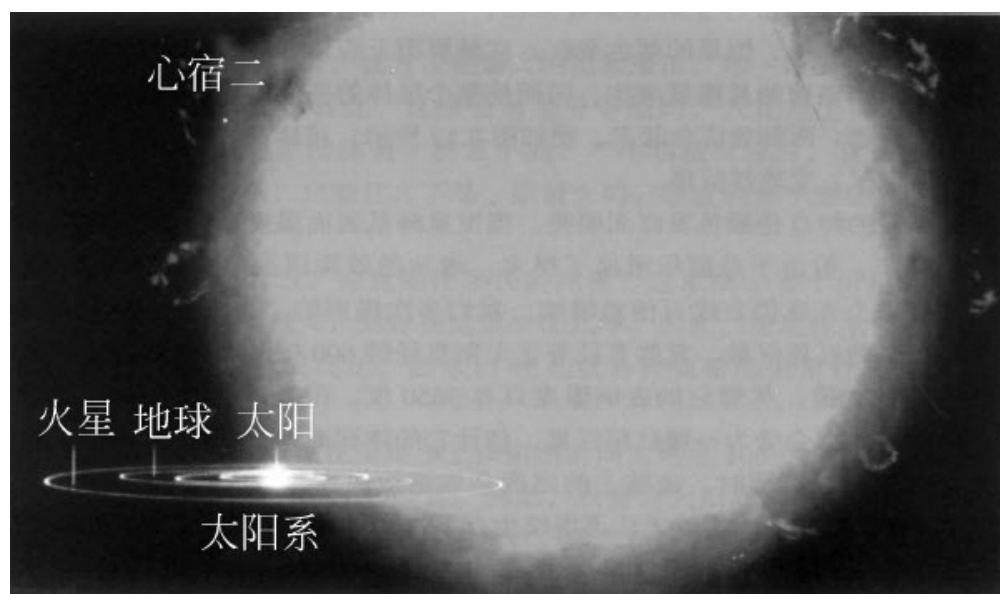


图2-7 红巨星心宿二与太阳系的比较

我们的太阳将来也会形成红巨星，大小与心宿二差不多。那时太阳中心部分的氢都已聚合成氦。热核反应扩展到外层，太阳的体积逐渐增大，温度逐渐降低，颜色变得越来越红。膨胀的太阳逐渐扩展到水星轨道，把水星吞到肚子里；再扩展到金星轨道处，把金星也吞进去；把地球上的江河湖海全烤干，然后把地球也吞进去，最后扩展到火星轨道处。

这时水星、金星和地球都在太阳的肚子里围着太阳的核心转。这是因为红巨星虽然温度达到4000度，体积很大，但除去核心部分之外，密度都极小，气体稀薄到比我们实验室所能制造的真空密度还稀，所以行

星能够长时间在红巨星的肚子里转，不会落向星体的中心。但那时地球上的生命肯定已荡然无存。

不过，大家不要担心。我们太阳在主序星阶段的寿命是100亿年，现在才过了50亿年，太阳正在年富力强的中年时期，演变成红巨星是50亿年以后的事情。未来的50亿年中，太阳将维持现在这种状态，所以大家尽可以放心地活着。

那么，50亿年以后怎么办呢？我们的子孙后代怎么办呢？没有关系。大家想一想，现代自然科学从哥白尼开始至今才500年，人类已经可以登月了，太阳演化到红巨星还需要50亿年的时间，那时的人类科学会有多么发达！

那时人类说不定早已移居到银河系中年轻的太阳系去生活了，甚至可能在地球上修个“喷嘴”，装好足够的燃料，再把绝大多数人类和生物“冷冻”起来，空气也冷冻到地球的表面上，只留下几个专业小组驾驶地球。“轰”的一下瞬间将地球像火箭一样发射出去，飞向事先考察确定的、适合人类生存的年青的太阳系，然后重新唤醒人类，继续在那里生活。

有人说，这简直是天方夜谭！我要说，这确实是幻想，但是是有道理的科学幻想。总之，50亿年后人类的命运用不着我们杞人忧天。

从红巨星到白矮星

前面谈到，红巨星的中部聚集着聚变反应后由氢生成的氦，外部的氢继续进行着聚变反应，生成的氦不断落向中部，使氦组成的中心部分质量不断增加，引力不断增加，温度不断增高，终于点燃了由氦聚合成碳的新的聚变反应。

这时红巨星的核心部分温度突然升高到1亿度左右，于是白矮星就形成了。白矮星周围的气体形成“行星状星云”（图2-8），这些星云逐渐扩散开去，最后只留下孤独的白矮星。

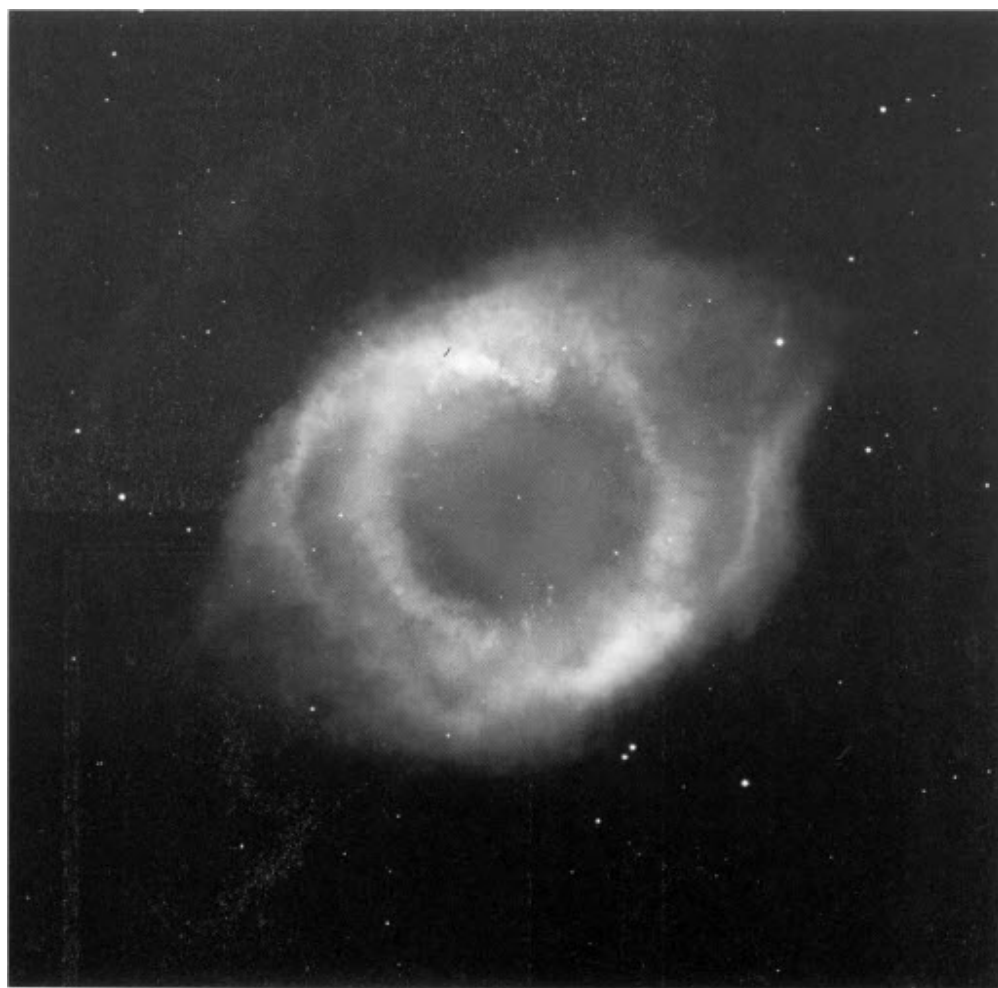


图2-8 行星状星云

大家知道，行星之所以不在万有引力作用下收缩为一个点，是因为原子靠得很近后，原子外部的电子云分布会发生变化，同种电荷互相靠近，静电斥力增加。万有引力越大，原子之间靠得越近，电子之间的静电斥力就越大。电磁力的排斥效应与万有引力的吸引效应相平衡，就使行星达到稳定的状态。

恒星的情况与行星不同，它们的温度很高，热排斥效应很强，这种热排斥与万有引力相抗衡，能够使恒星处于稳定状态。主序星和红巨星就是这种情况。

然而白矮星质量很大，密度很高，热排斥和电磁排斥效应都不足以抗衡它自身的万有引力。那么它靠什么来支撑呢？

研究表明，这时原子核外的电子壳层会被挤碎，使得电子能够在原子核形成的晶格内自由运动，或者说晶格似乎漂浮在电子的海洋中。这时电子间靠得很近，产生一种新的排斥效应——泡利不相容原理导致的排斥力。

泡利不相容原理是在研究原子结构时提出的。为了解释原子核外电子的排布，德国物理学家泡利提出如下原理：每个电子状态只能容纳一个电子。原子的每根电子轨道上有两个状态，所以每根轨道上只能存在两个电子。他把这一原理应用到玻尔的轨道模型上，给出了原子核外电子的壳层分布，成功地解释了元素周期律和光谱线。

后来的研究表明，一根轨道上的两个电子状态对应着电子的不同自旋。继而费米用统计理论阐明了半整数自旋的粒子（电子、中子、质子等）遵守泡利不相容原理的原因。

现在，这一原理又被用来解释白矮星物质中的强大斥力。上面已经说过，白矮星物质密度极大，电子壳层被挤碎，电子间靠得非常近。由于出现两个以上的电子挤占同一个状态的情况，于是相互间就会产生泡利斥力，这种力比热排斥和电磁排斥力都强。白矮星就是靠着这种泡利斥力的支撑，而不在强大的万有引力下塌缩的。

电子自旋的发现

首先想到同一轨道上的两个状态对应电子自旋的是泡利自己，但他很快又放弃了。他认为“自旋”这种概念来自经典物理，在量子理论中应该彻底放弃经典概念。所以放弃了电子有自旋的想法。

不久之后，美国青年物理学家克罗尼格产生了同样的想法，他跑去告诉泡利。泡利回答说：“你的想法很聪明，可惜大自然不喜欢它。”于是克罗尼格也放弃了自己的想法。

又过了一段时间，著名物理学家艾伦菲斯特的两个学生乌伦贝克和高斯密特也产生了类似的想法。艾伦菲斯特支持他们的想法，建议他们把稿子投给《自然》杂志，然后又建议他们去请教一下洛伦兹。

洛伦兹与他们都生活在荷兰。他们去后，洛伦兹说，我实在太忙，这样吧，你们把论文先放在我这里，我抽空看一下，你们过些日子再来。几天后，他们又去见洛伦兹，洛伦兹拿出一沓写满计算式的纸张交给他们，说：“恐怕有问题，如果电子有自旋，电子边沿的速度将超过光速，这违背相对论。”洛伦兹原先反对过爱因斯坦的相对论，现在态度转了 180° ，又赞同相对论了。

听了洛伦兹这番话，这两位觉得糟糕了！赶紧找老师艾伦菲斯特商量补救办法。老师建议他们赶快要求杂志社退稿，杂志社却答复：“十分抱歉，你们的稿子已经付印了，以后你们一定要注意，确定论文没有问题后再投稿。”为此，他们二人十分沮丧，老师安慰说，你们还年轻，发点错误文章没有关系，来日方长。

此后不久，玻尔到荷兰访问，艾伦菲斯特带着这两个学生到轮船码

头接玻尔。两位先生在前面走，他们二人帮玻尔提着皮箱，闷闷地在后面跟着。玻尔问艾伦菲斯特，你这两个学生好像有什么心事，不大高兴啊？艾伦菲斯特把他们投稿的事告诉了玻尔。

玻尔一听，说：“电子有自旋？这个想法很好啊！至于超光速的问题可以暂时先不管。”玻尔的称赞顿使乌伦贝克和高斯密特欣喜若狂。

这篇文章登出后，海森堡立刻写信给他们表示赞同，爱因斯坦也对他们的工作大加赞赏。

实际上，乌伦贝克和高斯密特的文章基本是正确的，当时出现超光速的问题是因为把电子的半径估计得太大了。事实上，电子非常小，至今还没有测出电子的半径。当然“自旋”这个概念，恐怕也确实不能像经典物理那样去理解。

文章发表前后，乌伦贝克和高斯密特的情绪经历了坐过山车一样的起伏，但他们终于没有错过“电子自旋”这一杰出的发现。

从白矮星到黑矮星

综上所述，白矮星靠着泡利斥力与万有引力相平衡，保持稳定状态。印度物理学家钱德拉塞卡的研究表明，质量小于 $1.4M_{\odot}$ 的白矮星将来会一直维持这种稳定状态，并逐渐冷却，形成黑矮星。

黑矮星实际上是由碳和少量的氧组成的巨大金刚石，在宇宙空间飘荡。谁要弄到一颗黑矮星，或者其上的一块物质，就足以发大财了。只可惜到今天为止，虽然发现了大量的白矮星（大约占恒星总数的十分之一），但还没有找到一颗黑矮星。这是因为白矮星冷却成黑矮星大约需要100亿年，我们宇宙的寿命至今才137亿年，可能一颗黑矮星都还没有形成。

白矮星的质量上限

然而，白矮星并不是所有恒星晚期的唯一归宿。印度青年物理学家钱德拉塞卡发现，残余质量超过1.4个太阳质量的白矮星状态的星体，由于万有引力过大，电子之间的泡利斥力将抵抗不住引力的吸引，星体将继续塌缩。

钱德拉塞卡有着极好的数学物理基础，他在研究中认识到，当万有引力迫使电子相互靠近时，电子运动速度将被迫加快，对于质量超过 $1.4M_{\odot}$ 的白矮星状态的星体，电子运动速度会接近光速，形成相对论性电子气，这时它们的泡利斥力会突然减弱，于是星体将不可能停留在白矮星状态，而会继续塌缩。

因此，钱德拉塞卡得出结论，白矮星存在一个质量上限—— $1.4M_{\odot}$ 。宇宙间不存在超过这个质量上限的白矮星。这个上限，后来被称为钱德拉塞卡极限。

钱德拉塞卡在印度大学毕业后，打算到英国去学习天体物理。他是在从印度到英国的旅行途中，在航船上完成这一发现的。

到了英国后，钱德拉塞卡与一些天体物理学家讨论了这一发现，后来又去征求著名天体物理学家爱丁顿的意见。他满以为爱丁顿会赞赏他的发现，没想到爱丁顿认为他的结论是错误的，让他好好检查一下自己的计算。此后他反复检查，都没有发现错误。然而几次找爱丁顿谈，都被爱丁顿不假思索地否定。

最后，爱丁顿看钱德拉塞卡一直坚持自己的结论，就对他说：不久

将在伦敦举行一次天体物理讨论会，你可以在会上报告一下你的工作，我可以运用我的影响，给你争取双倍的发言时间，让你充分讲解一下自己的工作。

开会的前一天，钱德拉塞卡与爱丁顿一起共进晚餐。钱德拉塞卡问爱丁顿：“您明天也有报告吗？”爱丁顿说：“有。”“什么名称呢？”“与你的报告名称一样。”钱德拉塞卡马上紧张起来，他觉得，爱丁顿是不是要抢夺自己的研究成果了。

第二天，轮到钱德拉塞卡作报告了，他向与会者散发了自己报告的预印本。他报告完之后，爱丁顿走上了讲台，他发言称：“钱德拉塞卡刚才的报告完全是错误的。”说着就把手中拿着的钱德拉塞卡的论文预印本撕碎。

爱丁顿认为，如果钱德拉塞卡的论文正确，白矮星将塌缩成一个点，一个物质密度和时空曲率都为无穷大的点，难道自然界存在这样的物理状态吗？谁看到过？所以，钱德拉塞卡的工作肯定是错误的。

由于爱丁顿的崇高威望，所有与会者都相信，钱德拉塞卡闹了个大笑话。散会时，一个朋友走到他身边，对他说：“钱德拉塞卡，这次糟透了，实在糟透了。”

其实，爱丁顿从未仔细阅读过钱德拉塞卡的论文，并没有发现论文的计算中真的有什么错误。他只是简单地分析了一下，就武断地否定了钱德拉塞卡的工作。爱丁顿之所以底气很足，还有一个原因，就是爱因斯坦赞同他的看法。

钱德拉塞卡24岁时做出这一发现，73岁时因为这一发现获得了诺贝尔物理学奖。钱德拉塞卡获奖后没几年就去世了。险些错过获奖的机

会，因为诺贝尔奖只颁发给活着的科学家。

不过，在获奖之前好多年，钱德拉塞卡的研究成果就已经得到了学术界的承认。在一次研讨会上，他走到泡利身边，请他看一下自己的论文。泡利拿过来一看，说：“这篇文章我看过。”他问：“您觉得我的文章怎么样？”泡利说：“很好啊。”钱德拉塞卡说：“爱丁顿教授说我的文章不满足您的不相容原理。”泡利说：“不，不，不。你的文章满足我的不相容原理，可能不满足爱丁顿不相容原理。”

泡利很少赞扬别人的工作，这次赞扬了钱德拉塞卡，但把爱丁顿讽刺了一下，可能他认为爱丁顿有种族歧视。应该说明的是，后来爱丁顿也承认了钱德拉塞卡的结论是正确的，并向他道了歉。

对中子星的预言

超过 $1.4M_{\odot}$ 的恒星是否就真的会塌缩成一个点呢？研究的结果并非如此。在万有引力的强大压力下，电子会被压入原子核中，与核里的质子“中和”成中子，形成一颗主要由中子构成的星——中子星。

这种星靠中子之间的泡利斥力来支撑。中子间的泡利斥力比电子间的泡利斥力大，可以支撑住质量不超过2~3个太阳质量的星体。这个质量上限称为奥本海默极限，是美国物理学家奥本海默首先给出的。

由于中子物质的物理状态研究起来很困难，目前还没有得到公认一致的物态方程，所以奥本海默极限的值不像钱德拉塞卡极限那么肯定，只能确定在 $2\sim 3M_{\odot}$ 范围内。超过这一极限的星体，中子间的泡利斥力顶不住万有引力，星体将塌缩成黑洞。

下面我们来介绍一下中子星的预言和发现。为此我们首先要介绍一下中子的发现。早在20世纪20年代，英国物理学家卢瑟福就怀疑原子核中是否存在一种质量与质子相似，但不带电的粒子。他是从元素原子序与原子量的关系作出这一猜测的。

例如氦元素，原子序为2，原子量为4。原子序2对应氦核中的两个质子，它们拥有2个单位的原子量，余下的2个单位原子量是什么贡献的呢？作出贡献的物质应该不带电，很像两个不带电的质子。卢瑟福猜测，这种未知粒子可能是电子和质子靠得非常近而形成的一种复合粒子。

1930年，普朗克的研究生波特在用 α 射线轰击铍时，发现了一种穿

透力很强的不带电的射线，他认为是 γ 射线（波长很短的电磁波）。

约里奥·居里夫妇（即居里夫人的大女儿和大女婿）对此进行了仔细研究，发现这种射线能够从石蜡中打出质子。他们也认为这是一种波长极短的 γ 射线。他们二人学化学出身，在研究中没有从物理的角度多加考虑。实际上，如果从能量守恒和动量守恒考虑，波长再短的 γ 射线也不可能从石蜡中打出质子。

卢瑟福的学生查德威克正在根据老师的猜测寻找中子，没有找到。这时他看见了约里奥夫妇的论文，真是喜出望外：“他们看见了中子还不知道。”于是他设计了类似的实验，报告自己发现了中子，先发一篇短文《中子可能存在》登在了《自然》杂志（*Nature*）上，接着发了一篇长文《中子是存在的》，登在英国皇家学会会报上。于是中子被发现了。

约里奥夫妇十分懊恼，他们错过了伟大的发现。正如法国著名微生物学家巴斯德所说：“机遇只钟情于有准备的头脑。”他们的头脑没有准备，没有中子这个概念。不过，他们没有停止不前，而是继续努力，很快发现了人工放射性，即用人工方法制出了放射性元素。此前发现的放射性元素都是天然的。

1935年，诺贝尔奖评委会一致认为，中子的发现应该获奖。有人主张应该由查德威克和约里奥夫妇分享这次的诺贝尔奖，但评委会主任是查德威克的老师卢瑟福。他说：“约里奥夫妇那么聪明，他们以后还会有机会的，这次的奖就给查德威克一个人吧。”于是，查德威克因为中子的发现获得了1935年的诺贝尔物理学奖。

当年下半年评化学奖，物理学奖与化学奖是由同一个委员会评，大

家一致同意把当年的诺贝尔化学奖授予约里奥夫妇，理由是“发现人工放射性”。其实，大家也在想：中子的发现他们也有一份功劳。

1954年，波特也因为宇宙线的研究获得了诺贝尔物理学奖，评委会大概也考虑了他对中子发现的贡献。这下皆大欢喜，4个对中子发现有贡献的人，都获得了诺贝尔奖，相当公平。

约里奥夫妇在发现人工放射性后，又首次观察到了核裂变，此后又首先提出，并在实验中观察到铀元素裂变的链式反应，打下了制造原子弹和核反应堆的理论基础，为人类利用原子能作出了重大贡献。

1932年，当发现中子的消息传到丹麦首都哥本哈根时，玻尔把理论物理研究所的人都聚集到一起，让大家畅谈一下发现中子的感想。当时，在那里访问的苏联青年物理学家朗道做了一个即席发言。他认为，宇宙中可能存在主要由中子构成的天体，这是对“中子星”最早的预言。此后，许多物理学家对这一问题进行了深入的研究。

脉冲星-中子星

然而中子星的发现则是30年之后的事情。1967年，英国天体物理学家休伊士和他的女研究生贝尔，在用射电天线对天空作巡天观测时，偶然发现了中子星。

我们看见的天体，都是发可见光的。但是有的天体只发无线电波，不发可见光。还有的既发无线电波又发可见光。来自空间的无线电波可以给我们提供许多信息，于是天文学家设计了一些天线装置，专门用来接收来自宇宙空间的无线电波。休伊士设计了一组天线阵，希望全面收集来自宇宙空间的无线电信号。为此他计划做“巡天观测”，即把天空划分成一个个天区，一个一个去扫描接收。他把这个任务交给了贝尔（图2-9）。



图2-9 休伊士、贝尔与他们的天线阵

夏天的一个夜晚，贝尔突然发觉接收的噪声背景中似乎隐藏着微弱的信号，她反复观测反复处理，终于发现确实存在着有规则的脉冲信号。她立刻打电话给自己的老师休伊士，休伊士赶来分析后，确认这是一种来自宇宙深处的规则脉冲。

他们原以为这是外星人在和我们联络，就给这一发现起了个代号叫“小绿人”。后来发现此脉冲信号非常规则，频率和振幅都毫不变化，不可能负载“外星人”的信息。而且不久之后又发现了几个这样的无线电波源，不可能有这么多有智慧的“小绿人”在联络我们。

后来弄清楚了，这是一种以前不了解的星体，它们发出频率严格不变的脉冲，于是称它们为脉冲星。再后来，人们终于弄清楚了，脉冲星就是中子星。

巨大的主序星塌缩成中子星时，恒星具有的磁场在星体塌缩过程中依然保持。但由于恒星体积大为缩小，原来主序星表面并不惊人的磁场在中子星表明显得异常强大，使得大量电子围绕磁力线转动，产生沿磁轴方向的电磁辐射。这种辐射非常强大，就像探照灯的光柱一样射向宇宙深处。

与普通恒星一样，中子星的磁轴与自转轴一般不重合，于是这些“光柱”就随着中子星的自转而在空间中扫描，每扫描地球一次，我们就接收到一个脉冲（图2-10）。

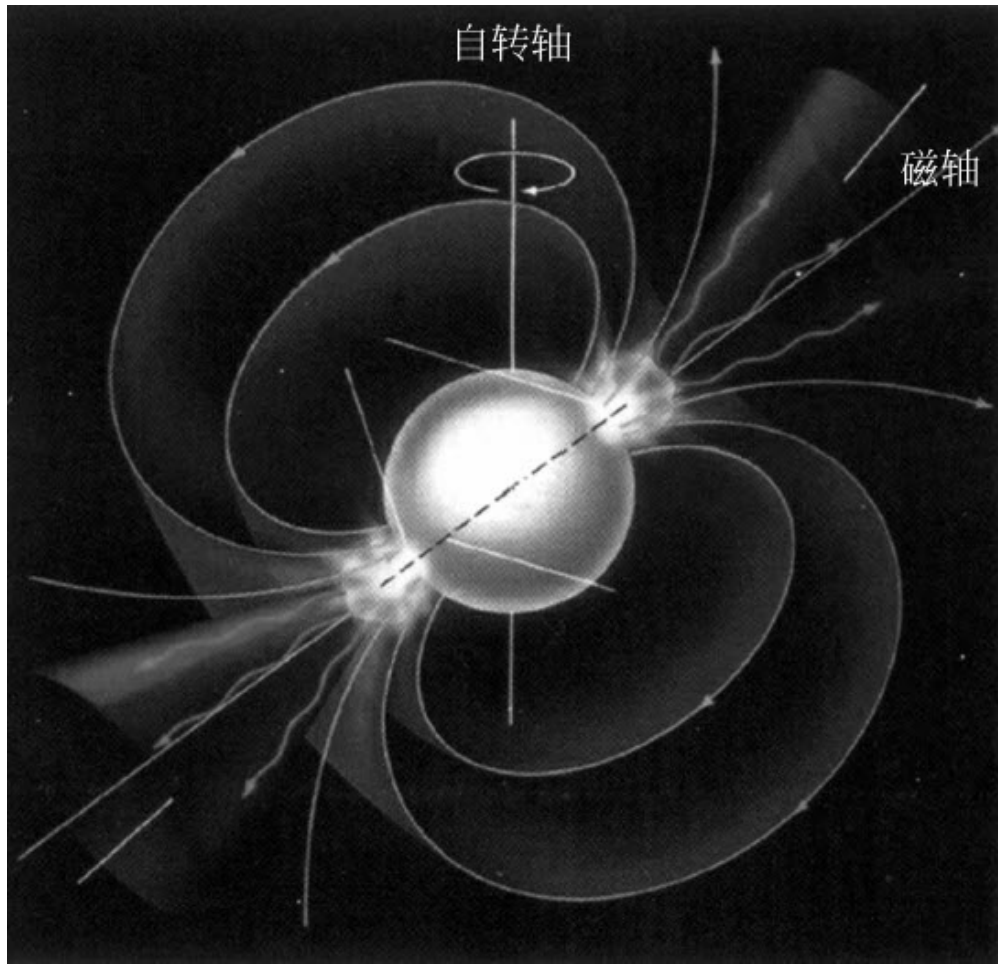


图2-10 脉冲星-中子星

主序星塌缩成中子星时，转动惯量大为减小，但角动量必须守恒，所以中子星的自转角速度极高，可达每秒几百圈，因此地球上观测到的、“光柱”扫描产生的脉冲频率很高。

这就是休伊士和贝尔发现的脉冲星，它们本质上是中子星。科学家终于发现了30多年前就已预言的中子星。

近年来的研究表明，中子星具有复杂的结构。它的主体部分由中子态物质组成，但会有大约百分之一的质子，以及数目相同的电子。

这是因为自由中子不稳定，会衰变成质子和电子，这百分之一的质

子，填满了中子可以衰变“抵达”的“基态”，使中子处于稳定状态，不再衰变。中子星外层有一个处于白矮星状态的铁壳，有人认为星体的中心部分还可能处在“夸克态”，好像是一锅“夸克汤”。

中子星表面温度极高，可达上千万度。由于引力巨大，中子星表面很平，最高的“山峰”也不过10cm。中子星的大气层很薄，也只有大约10cm。

不公平的颁奖

1974年，诺贝尔奖评委会为脉冲星的发现颁奖，奖金只给了休伊士一个人，忽略了贝尔。这件事在天体物理界引起了轩然大波，许多人为贝尔抱不平。休伊士出面说，他认为诺贝尔奖只颁给自己一个人是对的，因为天线阵是自己设计的，观测任务是自己布置给贝尔的，贝尔只不过是个执行者。

但许多专家认为，休伊士布置的任务只是一般的巡天观测，只是寻找众所熟知的那些射电源，并没有让贝尔寻找脉冲星。而脉冲星是一类原来未知的、非常特别的射电源，而且其信号极其微弱，如果不是贝尔细心观测，认真从噪声中提取信号，此发现很有可能会错过。

贝尔则表现得十分有涵养，从来没有批评过她的老师，也从来没有为自己申诉，这更加受到天体物理界学者们的同情。一些人对休伊士忿忿不平，霍金干脆在自己著作中叙述脉冲星发现的地方，只放了贝尔一个人的照片，并称她是脉冲星的发现者。

诺贝尔奖评委会对此事没有再说什么，但在以后的颁奖中注意了不应忽略学生的贡献。

超新星爆发

对于中子星的研究，我们中国人是有贡献的。那是在宋朝的时候。宋仁宗至和元年，中国的天文学家观察到一次超新星爆发（图2-11）。中国古代称超新星为“客星”，就是在本来观察不到恒星的天空位置，突然出现了一颗亮星，过了一段时间后，它又暗下去，逐渐消失。



图2-11 宋史中关于1054年超新星的记录

宋仁宗就是中国戏剧小说中“狸猫换太子”的那个太子。他在位的40多年，是北宋经济、文化最繁荣的时期。

至和元年，即公元1054年，中国人发现在“天关”附近出现了一颗客

星，有23天时间白天就能看见，像太白金星（即启明星，长庚星）那样亮。后来又有一二年时间晚上还能看见，再后来就消失了。日本、越南、朝鲜的古籍中也记录了这颗超新星，但没有指明方位，只有中国人准确地指出了它出现在天空的位置。

超新星，其实就是恒星的爆炸。超红巨星演化的晚期，聚集在星体中心的氦越来越多，先发生聚变反应，形成由碳和氧组成的白矮星状态的物质。由于这类恒星的质量巨大，一般超过8个太阳的质量，聚集在星体中心的碳和氧的数量超过钱德拉塞卡极限，电子的泡利斥力顶不住万有引力，所以星体会进一步收缩，温度达到30亿度左右，爆发碳、氧聚合成铁、硅等重元素的聚变反应。

随后，白矮星状态的铁核会进一步收缩升温，能在0.1秒的时间内升到50亿度，形成中子态的物质，这时星体外层的白矮星状态的铁壳层会突然塌缩，砸在中子态的核上，然后反弹出去，形成大爆炸。这就是外界观察到的超新星爆发。

超新星爆发把重元素崩向宇宙空间，其中许多被年轻的恒星吸引了过去，围绕这些“太阳”转，逐渐形成固体的行星。事实上，所有固体行星，包括我们的地球，都是超新星爆发的渣子堆积而成的。宇宙中原本只有氢、氦等轻元素，生成白矮星时开始出现碳和氧，超新星爆发过程才产生了铁、硅等重元素，并把它们抛向广阔的宇宙空间。

没有超新星爆发，就没有重元素，没有固体行星，当然也就不可能出现人类。超新星对于宇宙的演化、生命的出现和演化都是极为重要的。

除了为生命提供重元素，为人类提供“立足之地”外，超新星爆发还

会影响生物的进化。爆发时产生的强大辐射流，会使一些生物灭绝，同时为新物种的诞生和进化创造条件。地球上发生过若干次类似于恐龙灭绝这样的物种突变，造成这种突变的可能原因之一，就是在离我们太阳系较近的地方，发生了超新星爆发。

蟹状星云

1731年，英国的一位天文爱好者（一位医生），在金牛座中发现一个螃蟹状的星云（图2-12）。200年后一位西方天文学家和一位汉学家联手研究后，认识到这个星云就是1054年超新星爆发的遗迹。这里面传教士功不可没。明清时期，许多西方传教士来到中国，其中不乏有帝国主义的走狗，但大多数还是虔诚的基督徒，他们为宣传基督教来到中国，在传教的同时也为中国带来了西方的天文学和数学知识。



图2-12 蟹状星云

有些传教士在中国人的帮助下把有关的书籍翻译成中文，同时也把中国的历史文化介绍给西方，其中包括中国丰富的历史记录，于是包括关于“至和元年客星（即1054年超新星）”的许多天文记录传到了西方。传教士们的活动，极大地促进了中西方文化交流，使中国人大开眼界，也使西方人大开眼界。

1928年，天文学家哈勃测出了蟹状星云的膨胀速度，这个由气体和尘埃组成的星云，正在以1100千米/秒的速度扩散。星云的中心有一颗小暗星。1944年，人们认识到这颗小暗星和蟹状星云都是1054年超新星爆发的遗迹。从这个星云的膨胀速度不难反推出，组成星云的气体 and 尘埃正是1054年左右从那颗小暗星处喷出的。

中子星与黑洞的形成

更为惊人的是，1968年，天文学家们发现这颗小暗星是一颗脉冲星，人们终于认识到脉冲星（中子星）是超新星爆发的产物。如果没有中国古代的有关记录，得到这一结论还不知道要等到什么时候。

超新星爆发并不常见，我们银河系大约每100年爆发4颗。超新星爆发的规模非常大，超新星在一天里发出的光，相当于太阳在一亿年内发出的光。一颗超新星的亮度往往可以比一个星系（一般有上千万颗恒星）发的光还亮，从图2-13就可以看出。超新星与整个星系亮度的比较。图（a）是超新星爆发前的情况，图（b）是超新星爆发时的情况。

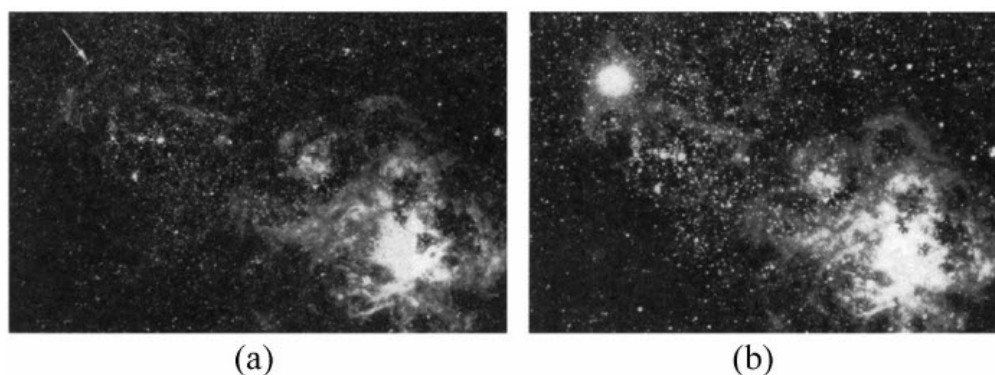


图2-13 超新星与星系亮度的比较

少年时代的我第一次听说超新星爆发时，不少天没有睡好觉。担心太阳如果超新星爆发怎么办，人类不就灭亡了吗？后来看到一些书上讲，太阳大小的恒星不会超新星爆发，才安下心来。真是杞人忧天啊！科学告诉我们，只有超过太阳质量8倍的主序星，演化到晚期时，才可能发生超新星爆发，我们的太阳是绝对安全的。

研究表明，除去形成中子星外，超新星爆发还可能形成黑洞。爆发时残留质量超过奥本海默极限的星体，中子间的泡利斥力抵抗不住万有

引力，星体将塌缩到它的引力半径 r_g 之内，即事件视界之内，形成黑洞。

I a型超新星和宇宙的加速膨胀

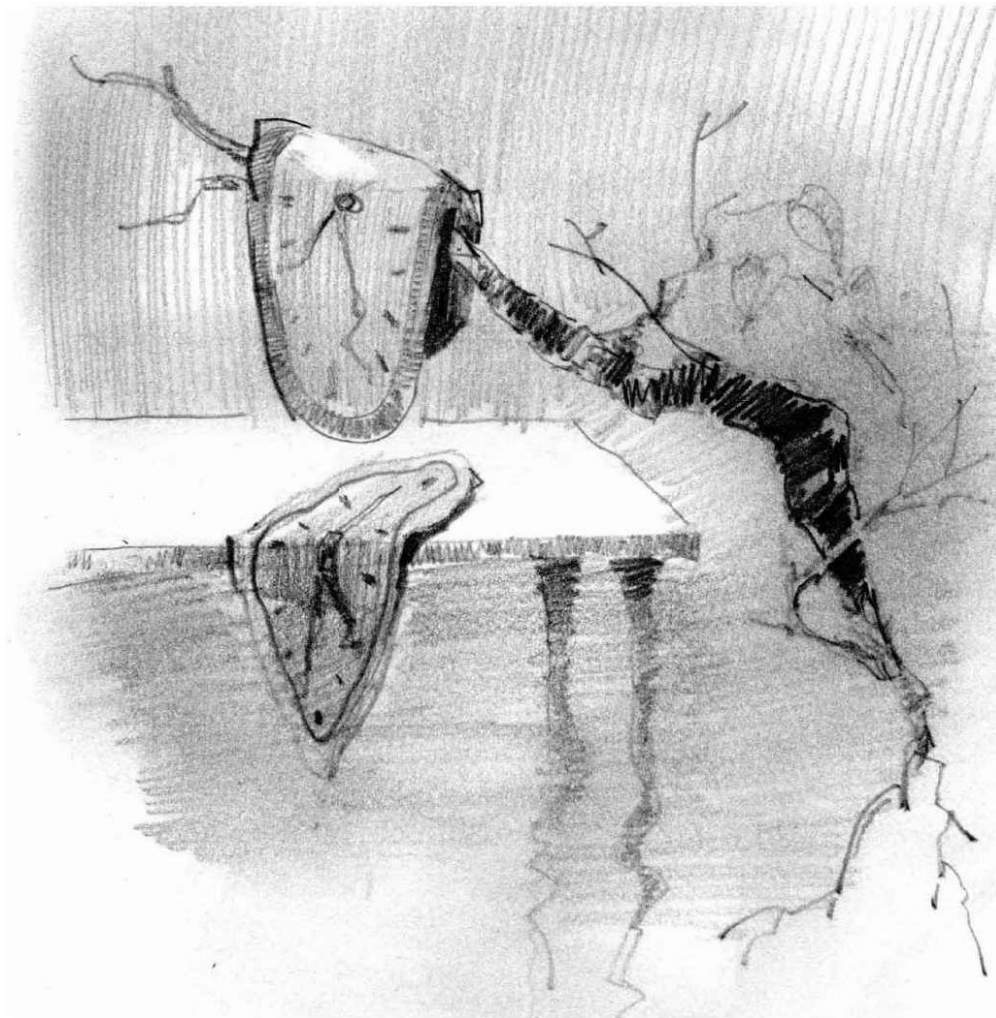
此外，还有一种超新星，即 I a型超新星，它爆发时全部炸光，不留残骸。这是一种特殊的星体，它原本是一颗白矮星，质量小于钱德拉塞卡极限 $1.4M_{\odot}$ ，本来不会超新星爆发。

但是这种白矮星处在一个双星系（由两颗恒星组成的太阳系）或一个聚星系（由多颗恒星组成的太阳系）中，它吸积其他恒星的气体，形成吸积盘。气流物质在吸积盘上边旋转边落入白矮星，使它的质量逐渐增加。当白矮星的质量超过 $1.4M_{\odot}$ 时，电子间的泡利斥力抵抗不住万有引力，于是这颗星塌缩并爆炸了。由于它的质量不足以形成中子星或黑洞，于是全部炸飞，不留残骸。

I a型超新星的研究使科学家重新确认了远方星系离我们的距离，认识到我们的宇宙在60亿年前从减速膨胀变成了加速膨胀，从而造成了世人皆知的暗能量疑难。

暗能量是一种在宇宙空间均匀分布的、透明的、压强为负的物质，它在宇宙中的密度不发生变化，使得宇宙在膨胀过程中内部的暗能量越来越多，产生的负压强越来越大，终于排斥效应超过了万有引力的吸引效应，从而使宇宙从减速膨胀变成了加速膨胀。通过前面的讨论，我们从天文学的角度，从恒星演化的角度看到了黑洞形成的可能性。目前白矮星与中子星都已经发现了。大小与中子星差不多大的黑洞似乎不应该不存在。

下面，我们将把讨论的注意力，集中到本书的主题黑洞上。为此，我们先介绍一下黑洞的理论基础——广义相对论。



绘画: 张京

第三章 弯曲的时空

爱因斯坦1915年完成的广义相对论是“关于时间、空间和引力的理论”。这一理论可以看做万有引力定律的发展，它认为万有引力不是一般意义上的力，而是一种几何效应，是时空弯曲的表现。

狭义相对论的成就

1905年，爱因斯坦在“相对性原理”和“光速不变原理”的基础上导出“洛伦兹变换”，建立起相对论（即今天所说的狭义相对论）的大厦。他给出了惯性系中“动尺缩短”、“动钟变慢”、“质能关系 $E=mc^2$ ”、“双生子佯谬”等重要而新奇的结论。相对论突破了牛顿理论的框架，展现出全新的物理体系和全新的时空观。

爱因斯坦指出，自己的相对论与牛顿的经典物理学的关键差别不在“相对性原理”，而在“光速不变原理”。因为伽利略早就正确地阐述了相对性原理，牛顿在自己的力学中也应用了这一原理。只是洛伦兹等人为了解释迈克耳孙实验，对相对性原理产生了怀疑，把水搞得有点浑。爱因斯坦说，我本人只是坚持了这一原理，并无特别的创新。第三章弯曲的时空看不见的星：黑洞与时间之河

爱因斯坦认为，自己最大的突破是认识到光速是绝对的，真空中的光速不仅在同一惯性系中是均匀各向同性的，而且与观测者相对于光源的运动速度无关。

上句话的前一半说“光速在同一惯性系中均匀各向同性”，这是一个“约定”，即“规定”。只有做了这一“约定”，才可以校准不同地点的钟，从而可以在全空间定义统一的时间（这一点我们将在第九章作详细讨论）。应该说，这一约定是建立相对论的前提。

“光速不变原理”则是指上句话的后一半，“光速与观测者相对于光源的运动速度无关”。这一原理的意思是，相对于光源静止的观测者，迎着光束以速度 v_1 相对于光源运动的观测者，以及顺着光的前进方向，

以速度 v_2 远离光源的观测者，测到的真空中的光速都是同一个 c 值。这确实是让人难以理解的。

而且，承认“光速不变原理”，就意味着必须承认“同时”这个观念不再是绝对的，而成了相对的。这就是说，在高速行驶的火车上，车上的人认为车头与车尾“同时”发生的两件事，在静止于地面上的观测者看来，不再是同时发生的。当然，由于火车速度不够高，在日常生活中这一效应看不出来，但是如果火车速度接近光速，这一效应将十分明显。

理解“光速的绝对性”，及其导致的“同时的相对性”，是十分困难的，这个难题曾经困扰了爱因斯坦很长时间，大约在一年以上。他一想通这点，所有的问题就迎刃而解了，“动钟变慢”、“动尺收缩”等相对论效应就都自然得出了。

洛伦兹与庞加莱都曾十分接近相对论的发现，但他们都没有认识到“光速不变原理”，没有谈论过、也根本没有想到过“同时”这个概念会是相对的，从而与相对论的发现失之交臂。

洛伦兹与庞加莱为相对论的发现做了许多奠基性的准备，但他们都因为没有认识到“光速不变原理”，而没能跨入相对论的大门。所以爱因斯坦是相对论的唯一创建人。

狭义相对论的困难

正当全世界为相对论的成功而欢欣鼓舞时，爱因斯坦本人却冷静地看到了自己的理论存在严重缺陷。

首先，作为“相对论”基础的惯性系，现在无法定义了。牛顿认为，存在绝对空间，所有相对于绝对空间静止和作匀速直线运动的参考系都是惯性系。爱因斯坦的相对论认为不存在绝对空间，牛顿定义惯性系的方法显然不适用了。

有人建议，把惯性系定义为，不受力的物体在其中保持静止或匀速直线运动状态的参考系，也即把牛顿第一定律（即惯性定律）视作惯性系的定义。

但是，什么叫不受力呢？有人以为，一个物体不与其他物体接触，就肯定不受力。这种想法不对，因为自然界中还存在看不见摸不着的各种场，例如，引力场、电磁场等，它们都可能对物体“施力”。

也许有人会说，如果物体在惯性系中，保持静止或匀速直线运动的状态，就可以定义为“不受力”。读者一下就会看出，这里存在逻辑上的循环。定义“惯性系”要用到“不受力”。定义“不受力”，又要用到“惯性系”。这样的定义方式，在物理学中是不可接受的。

爱因斯坦注意到的另一个缺陷是，万有引力定律写不成相对论的形式。有几年，爱因斯坦致力于把万有引力定律纳入相对论的框架，几经失败后，他终于认识到，相对论容纳不了万有引力定律。

这两个缺陷非常严重。他的相对论是研究惯性系之间的关系的，也

就是说，相对论是建立在惯性系的基础上的。现在，这个“基础”居然无法定义！另一方面，当时已知的力只有电磁力和万有引力两种，竟然其中的一种就放不进相对论的框架中，真是太令人遗憾了！

走向新理论的第一步：广义相对性原理

爱因斯坦反复考虑狭义相对论遇到的两个基本困难：①惯性系无法定义；②万有引力定律纳不进相对论的框架。他想，既然惯性系无法定义，不如就抛开惯性系，把自己的理论建立在“任意参考系（包括非惯性系）”的基础之上；把原来的相对性原理：“物理规律在一切惯性系中都相同”推广为“物理规律在一切参考系中都相同”。他把后者称为广义相对性原理，以别于原来的相对性原理。这样做确实躲开了定义惯性系的困难，但又产生了新的困难：非惯性系与惯性系不同，它有惯性力存在。如何处理惯性力呢？爱因斯坦想到了牛顿的水桶实验，以及牛顿和马赫对惯性力起源的不同看法。



在布拉格, 1912 年
图3-1 研究广义相对论时的爱因斯坦

牛顿的水桶实验

牛顿为了论证绝对空间的存在，曾设计了一个著名的思想实验：水桶实验。

一个装有水的桶，最初桶和水都静止，水面是平的（图3-2（a））。然后让桶以角速度 ω 转动，刚开始时，水未被桶带动，这时候，桶转水不转，水面仍是平的（图（b））。不久，水渐渐被桶带动而旋转，直到与桶一起以角速度 ω 转动，此时水面呈凹形（图（c））。然后，让桶突然静止，水仍以角速度 ω 转动，水面仍是凹形的（图（d））。

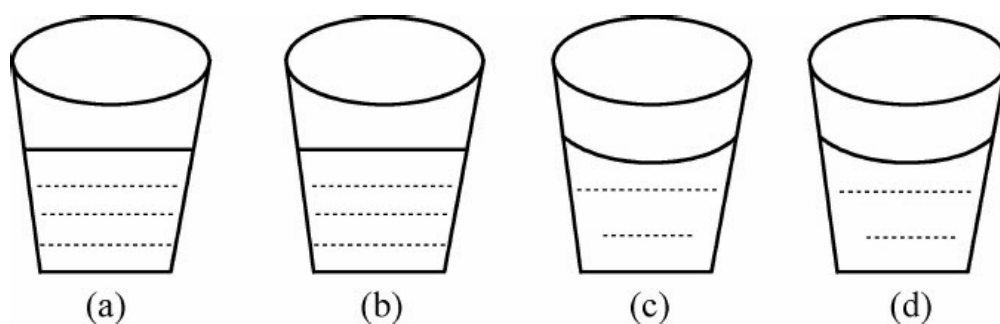


图3-2 水桶实验

（a）桶与水都静止；（b）桶转水未转；（c）水与桶一起转；（d）桶静止水转

在情况（a）和情况（c）中，水相对于桶都静止，但水面在情况（a）时是平的，在情况（c）时是凹的。而在情况（b）和情况（d）中，水相对于桶都转动，但水面在情况（b）时是平的，在情况（d）时是凹的。

显然，水面的形状与水和桶的相对转动无关。水面呈凹形是由于受到惯性离心力的结果。惯性离心力的出现既然与水相对于桶的转动无关，那么与什么有关呢？牛顿认为，与绝对空间有关。惯性离心力产生

于水对绝对空间的转动。牛顿认为，转动是绝对的，只有相对于绝对空间的转动才是真转动，才会产生惯性离心力。推而广之，加速运动是绝对的，只有相对于绝对空间的加速才是真加速，才会受到惯性力！通过水桶实验，牛顿论证了绝对空间的存在。

走向新理论的第二步：马赫原理

奥地利学者马赫只是一个三流的物理学家，他对物理学的具体贡献不大，主要是在流体力学中提出了“马赫数”。但是他敢于向“祖师爷”挑战，敢于批判牛顿，说牛顿的时空观不对。马赫的观点对青年爱因斯坦产生了深远的影响。

马赫认为根本就不存在什么绝对空间，一切运动都是相对的。他认为牛顿对惯性力起源的解释是错误的。马赫认为惯性力起源于物体间的相对加速，起源于作相对加速的物体之间的相互作用。

我们通常所说的受到惯性力的加速物体，是由于它相对于宇宙中的所有物质加速，这相当于该物体不动，整个宇宙的物质相对于它作反向加速。全宇宙的物质通过这种加速共同对该物体施加了“作用”，这种“作用”就是惯性力。

水桶实验中受到惯性离心力的水，是由于它相对于整个宇宙中的物质转动，这相当于水不动，全宇宙的物质相对于水反向转动，在这种相对转动中，全宇宙的物质都对水施加了“作用”，这一作用的表现就是惯性离心力。如果水只相对于桶转动，而相对于宇宙中的其他物质不转，虽然桶对水也施加了影响，但桶的质量与整个宇宙的质量相比微乎其微，转动的桶对水的“作用”也就微乎其微，所以这时水不会受到惯性离心力。

马赫这种认为惯性效应起源于物体之间的相对加速，从而起源于物质间的相互作用的思想，被爱因斯坦称为马赫原理。马赫原理并没有严格的物理陈述，更没有数学表达式，它只是一种定性的物理思想。

正是马赫的这一思想，给了爱因斯坦重要的启示，引导爱因斯坦提出等效原理，并进而建立起广义相对论的大厦。在广义相对论取得巨大成功之后，爱因斯坦高度评价马赫的这一思想，认为自己的广义相对论具体体现了马赫原理预期的效应。

不过，后来的深入研究表明，广义相对论与马赫原理并不一致。有趣的是，当时马赫还活着。马赫看到了狭义相对论，但据说没有看到广义相对论。马赫断然否认自己的思想与相对论一致，明确反对爱因斯坦的相对论。这一点令爱因斯坦十分遗憾。

然而，有一点可以肯定，马赫认为“不存在以太和绝对空间，一切运动都是相对的”，这一思想引导爱因斯坦走上了创立狭义相对论的正确道路；马赫认为“惯性效应起源于物质间相对加速产生的相互作用”的思想，又导致爱因斯坦猜测惯性力可能与万有引力有相同或相似根源，都起源于物质间的相互作用，从而引导爱因斯坦走上了创立广义相对论的正确道路，所以，马赫原理在历史上的贡献是应该肯定的。

引力质量与惯性质量

此外，爱因斯坦注意到惯性力的一个重要特点：惯性力与物体的惯性质量成正比。这个特点与万有引力非常相似，万有引力与物体的引力质量成正比。

在《自然哲学之数学原理》一书中，牛顿把质量定义为物体所含物质的多少。他说，质量就是物质的量，它等于密度和体积的乘积，并与物体的重量成正比。这样定义的质量，称为引力质量（ m_g ）。在这本书的另一处，牛顿又谈到物体的质量与它的惯性成正比。也就是说，他认为，使不同物体在外力作用下产生相同加速度时，物体的质量与所施外力的大小成正比。所以，他认为质量可以看做物体惯性的量度，这样定义的质量，称为惯性质量（ m_I ）。

牛顿认为，没有理由相信引力质量和惯性质量是同一个东西。但是，自由落体定律似乎告诉我们二者相等。自由落体运动是物体在万有引力作用下，依照牛顿第二定律所作的加速运动。万有引力定律可以写成

$$F = G \frac{Mm_g}{r^2} = m_g g \quad (3.1)$$

其中， r 、 M 分别为地球的半径和引力质量， m_g 为下落物体的引力质量， g 为引力场强。牛顿第二定律是

$$F = m_I a \quad (3.2)$$

m_I 为下落物体的惯性质量。二式联立可得

$$m_g g = m_I a \quad (3.3)$$

自由落体定律告诉我们。任何物体，不管其化学成分和重量，它们的下落加速度 a 都相同，都等于 g ，这就导致引力质量与惯性质量相等

$$m_g = m_I \quad (3.4)$$

不过，在牛顿时代这个实验相当粗糙，精度不高。于是，牛顿用单摆实验来做进一步的检测。单摆运动是摆球在重力和摆绳弹力共同作用下，按牛顿第二定律所作的运动。

单摆运动方程的一边是摆球所受的力，它依赖于摆球的引力质量 m_g ；方程的另一侧表示单摆在重力下的加速，它依赖于摆球的惯性质量 m_I 。这样算得的单摆公式为

$$T = 2\pi \sqrt{\frac{m_I l}{m_g g}} \quad (3.5)$$

如果这两个质量对于不同的物体有差异， m_I/m_g 对各种物体不是同一个常数的话，单摆运动的周期，对不同物体就会有所不同。

但牛顿没有观测到这种不同。他在千分之一的精度范围内证明了引力质量等于惯性质量。所以，人们都把单摆周期公式写为

$$T = 2\pi \sqrt{\frac{l}{g}} \quad (3.6)$$

爱因斯坦那个时代有个匈牙利物理学家Eötvös，中国人翻译成厄缶，他用扭摆实验在 10^{-8} 的精度之内没有查到引力质量和惯性质量的差异。相对论发表以后，迪克做到 10^{-11} ，俄罗斯的布拉金斯基做到 10^{-12} ，

都严格地证明了引力质量和惯性质量相等。在爱因斯坦那个时代，精度最高的是厄缶那个实验。爱因斯坦研究引力理论时知道这个实验。

爱因斯坦注意到：自由落体实验、牛顿的单摆实验和厄缶的扭摆实验都表明引力质量与惯性质量精确相等。他还想起了自己一直钦佩的物理学家兼哲学家马赫的见解：惯性力与万有引力相似，都起源于物体间的相互作用。

爱因斯坦终于认识到，“惯性”问题应该和“引力”问题合在一起解决，狭义相对论所遇到的两个困难实际上是同一个困难！

走向新理论的第三步：等效原理

这时，爱因斯坦不仅认识到引力与惯性力的相似性，而且认识到“引力质量与惯性质量相等”是经过严格实验检验的结论。经过反复思考后，他决定把这一结论往前推进一步，提出等效原理：惯性场与引力场局域等效。这就是说，在无穷小时空范围内，人们无法区分引力与惯性力。

爱因斯坦关于升降机（电梯）的思想实验（图3-3），最清楚地表达了他的等效原理思想。设想一个观测者处在一个封闭的升降机内，得不到升降机外部的任何信息。当他看到机内的一切物体都自由下落，下落加速度 a 与物体的大小及物质组成无关时（此时，他自己也感受到重力 Ma ， M 是他自身的引力质量），他无法断定自己处在下列两种情况的哪一种：

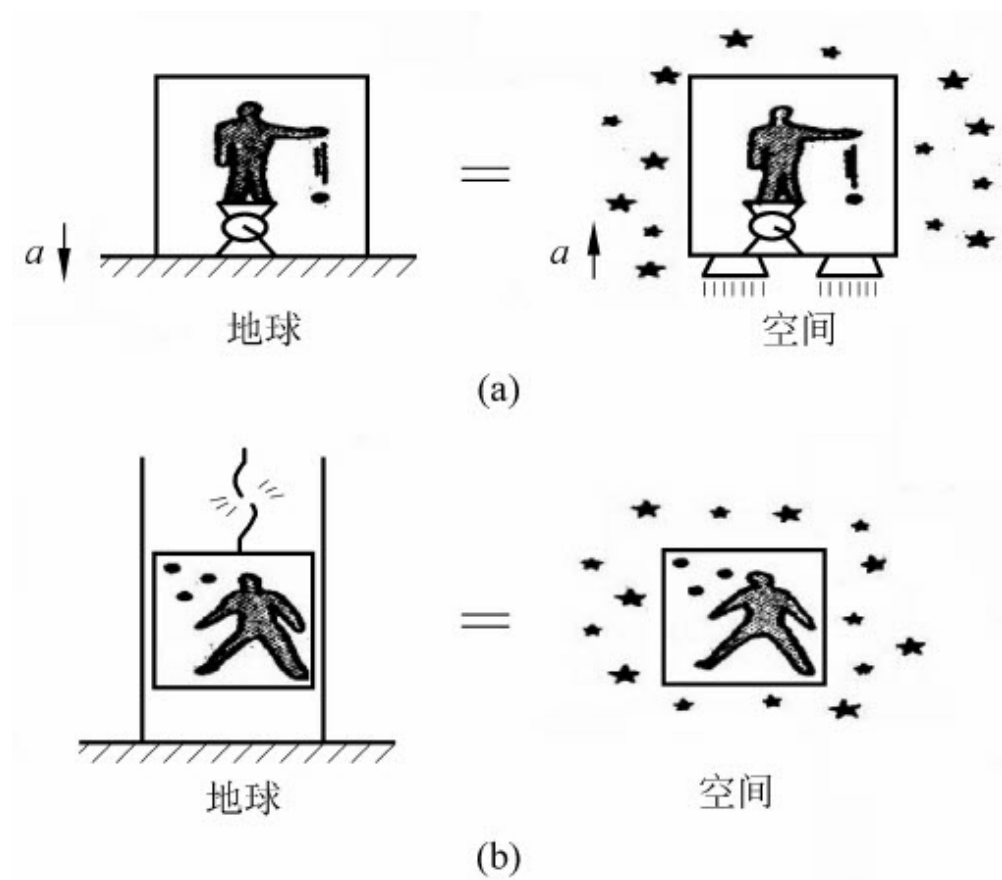


图3-3 爱因斯坦升降机

(a) 引力与加速：左边表示升降机在星球（例如地球）表面静止，右边表示升降机在太空中加速；(b) 自由下落与失重：左边表示升降机在星球引力场中自由下落，右边表示升降机在太空中作惯性运动

(1) 升降机静止在一个引力场强为 a 的星球的表面。

(2) 升降机在无引力场的太空中以加速度 a 运动。

当观测者感到自己和升降机内的一切物体都处于失重状态时，他同样无法断定自己处在下列两种情况的哪一种：

(1) 升降机在引力场中自由下落。

(2) 升降机在无引力的太空中作惯性运动。

造成上述现象的原因是，无法用任何物理实验来区分引力场和惯性场。即等效原理造成了上述不可区分性。

然而，引力场与惯性场还是有不同之处，它们在有限大小的时空范围内并不等效。例如，由于星球是球体，静置于星球表面的升降机，其内部的引力线有向星球中心汇聚的趋势，而在星际空间加速的升降机，其内部的惯性力线则是平行的。只要升降机不是无穷小，探测这些力线的灵敏仪器就可以区分这两种情况。

所以等效原理是一个局域性原理。也就是说，引力场与惯性场仅在无穷小时空范围内不可区分。

等效原理、马赫原理和广义相对性原理，形成了爱因斯坦新理论的物理基础。

惯性系的新定义

等效原理告诉我们，引力场中一个自由下落的、无自转的无穷小参考系，等价于在无引力场太空中作惯性运动的无穷小参考系。

如果我们把狭义相对论在其中成立的参考系定义为惯性系，那么，在无引力场太空中静止或作匀速直线运动的参考系和在引力场中自由下落且无自转的无穷小参考系，就都是惯性系（自转会引起惯性离心力和科里奥利力，从而偏离惯性系）。

然而，这两种惯性系实际上都不能严格实现，我们既不能使真实的参考系“无穷小”，又无法找到完全不存在引力场的太空区域，所以，在实际生活和科学实验中，只能找到近似的惯性系。

意大利物理学家费米是一位“多面手”，既能研究理论，又能搞实验。他的成就遍及统计物理、核物理、量子理论和广义相对论，还培养了6个得诺贝尔奖的学生。他曾经证明，一个无穷小的、无自转的自由下落参考系，可以在下落过程中，一直保持是一个惯性系。这种无穷小惯性系在时空中的运动，被称为费米移动。

猜想：引力是几何效应

等效原理还进一步告诉我们，当只有引力场与惯性场存在时，任何质点，不论质量大小，在时空中都会描出同样的曲线，自由落体实验已表明了这一点。再如，在真空中斜抛金球、铁球和木球，只要抛射的初速和倾角相同，这三个球都将在空间描出相同的轨迹。

这就是说，质点在纯引力和惯性力作用下的运动，与它的质量和成分无关。于是，爱因斯坦做出了物理思想上的又一个重大突破，他大胆猜测，引力效应可能是一种几何效应。万有引力不是一般的力，而是时空弯曲的表现。由于引力起源于质量，他认为时空弯曲起源于物质的存在和运动。

如何把时空几何与运动物质联系起来呢？爱因斯坦在建立新理论的过程中感到自己的数学知识欠缺，他需要新的数学工具。于是，他求助于自己的大学同学格罗斯曼。那时格罗斯曼已是大学的数学教授，他查阅了一批文献，然后告诉爱因斯坦，当时一些意大利人正在研究的黎曼几何和张量分析，也许对他有用。

奥林匹亚科学院

爱因斯坦早就对黎曼几何有一点定性的了解。他在专利局工作期间（1902—1905年），曾与几位热爱科学与哲学的好友组织了一个叫做“奥林匹亚科学院”的小组（图3-4）。

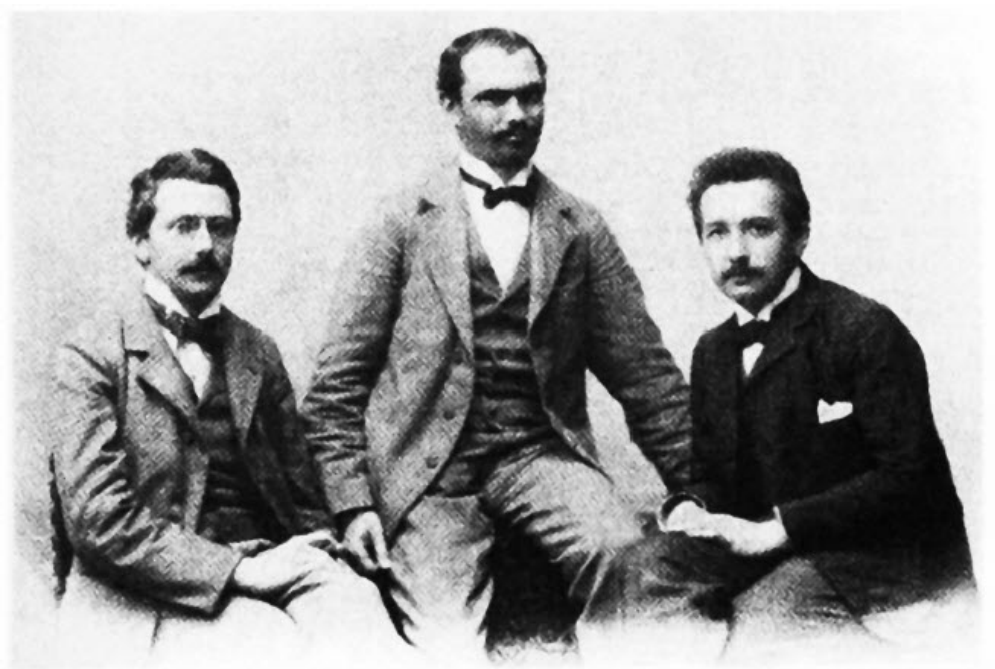


图3-4 爱因斯坦与奥林匹亚科学院中的朋友

与“奥林匹亚科学院”的哈比希特（左）和索洛文，约1902年

这是一个自由读书与自由探讨的俱乐部。小组的成员都具有大学文化水平，他们工作单位不同，专业背景也不同，有学物理的，有学哲学的，还有学工程技术的。这几个年轻人利用休息日或下班时间，一边阅读一边讨论，内容海阔天空，以哲学为主（特别是与物理有关的哲学），也包括物理、数学和文学。

他们充满热情地阅读、讨论了许多书籍，其中包括马赫的《力学史评》，这本书对牛顿的绝对时空观进行了猛烈的批判。马赫的见解深刻

地影响着年轻的爱因斯坦，他不惧权威的批判精神鼓舞着年轻的爱因斯坦。爱因斯坦曾多次强调，他提出狭义与广义相对论都与马赫的影响有关。

他们还阅读讨论了庞加莱的名著《科学与假设》，在此书中庞加莱简单提到过黎曼几何。这本书曾引起爱因斯坦和他的朋友们的极大兴趣。现在发现这本书中的内容居然有可能在自己的研究中派上用场，爱因斯坦十分高兴，于是他愉快地接受了格罗斯曼的忠告，开始关注和学习黎曼几何。

非欧几何的创立

下面我们简单介绍一下黎曼几何的建立。

欧氏几何，以它逻辑的严密，形式的完备和优美，两千年来为数学家和哲学家所倾倒。使人感到美中不足的是它的第五公设，即平行公理。此公理说，“过直线外的一点，可以引一条、并且只能引一条直线与原直线平行（不相交）”。

与其他公设比较，这个公设显得过长、过于复杂。人们自然希望第五公设能从其他公设推出，从而不再是一个公设。这方面的尝试开始于公元5世纪。一千多年中，许多杰出数学家为它绞尽脑汁，结果都一无所获。

无数前人的失败，终于使后人悟出了道理。第一个察觉其中奥妙的人大概是高斯。然而，由于欧氏几何在数学、哲学和神学中的神圣地位，高斯缺乏公开挑战的勇气，没有发表自己的观点。

最先提出并建立完整的新几何学的是俄国数学家罗巴切夫斯基（1792—1856年）。他是喀山大学的教授。他用“过直线外一点，可以引两条以上直线与原直线平行（不相交）”的新公设来取代第五公设。然而，他的理论在国内无人能懂，多次投稿均被拒绝，俄国彼得堡科学院甚至认为，“罗巴切夫斯基先生这方面的工作谬误连篇，今后不必理睬。”

罗巴切夫斯基只好把论文发表在喀山大学的学报上，外界几乎无人知晓。后来他出国演讲，宣传自己的新几何，但还是遭到冷遇。唯一听懂了他的理论的高斯，未敢公开表示赞同。高斯在日记和给友人的信中

写到，会场上，自己大概是唯一听懂了罗巴切夫斯基工作的人。

应该说明，高斯虽然未对新几何表态，却高度评价了罗巴切夫斯基的其他工作。经高斯提名，德国科学院授予罗巴切夫斯基通讯院士的光荣称号。然而，罗巴切夫斯基最杰出的工作，却长期得不到承认。他晚年双目失明，处境凄凉，但仍在学生的协助下，顽强地通过口述完成了自己的工作，并在逝世前，终于得到世界的认可。罗巴切夫斯基的新几何，被称为罗氏几何。

比罗巴切夫斯基稍晚一点，年轻的匈牙利数学家鲍耶（1802—1860年）也独立地研究了这个问题，并提出了与罗氏一样的新几何构想。鲍耶采用了反证法，企图从“第五公设不成立”引出谬误。他假设过直线外的一点可以引两条以上的平行线，试图推出谬误。然而他在反证的路上越走越远，却始终不见“谬误”的影子。鲍耶的思想突然产生了飞跃，认识到可以引入不同于第五公设的其他公理，例如把“过直线外一点可以引两条以上的平行线”作为公理，取代“第五公设”建立新的几何学。

鲍耶的父亲是高斯的同学，他把儿子的工作告诉高斯，征求他的意见。高斯回信说：我实在无法赞扬你的儿子，因为赞扬他就等于赞扬我自己，我早就得到过与他相同的结果。

鲍耶误解了高斯对他工作的评价，以为高斯要借其所处的地位来窃取他的成果，于是愤而终止了自己的研究。幸亏鲍耶的父亲把儿子的研究成果作为附录，收在自己的一本书中出版，否则，鲍耶的成就将无人知晓。

德国数学家黎曼用另一个公设来代替欧几里得的第五公设。他提出，“过直线外一点的任何直线都必定与原直线相交”，也就是说，一条

平行线也做不出来。他所建立的几何称为黎氏几何。

实际上，欧氏几何、罗氏几何、黎氏几何描述的是不同曲率的空间（图3-5）。欧氏几何描述零曲率空间（如平面），黎氏几何描述正曲率空间（如球面），罗氏几何描述负曲率空间（如伪球面、马鞍面）。弯曲空间中没有直线。罗巴切夫斯基等人谈论的直线实际是“短程线”，即两点之间的最短线，例如球面上的大圆周。平直空间中的短程线就是直线，短程线可以看作直线在弯曲空间的推广。罗巴切夫斯基等人所说的平行直线，实际上是“不相交的短程线”。

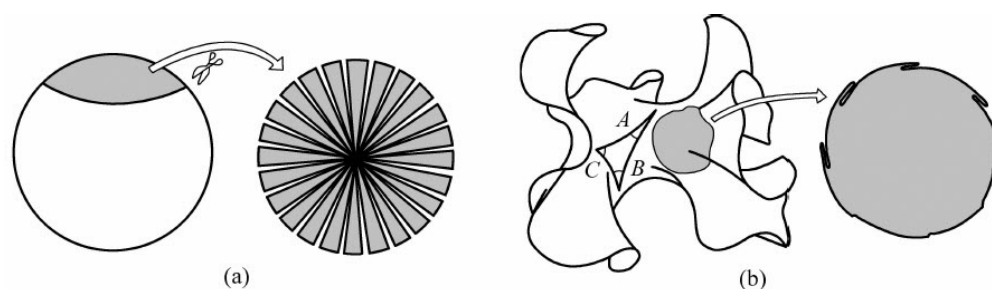


图3-5 正负曲率的二维空间

罗氏几何和黎氏几何，统称非欧几何。黎曼又把欧氏几何与非欧几何合在一起，成为黎曼几何。

黎曼几何简介

下面我们以球面上的几何为例，来解释一下黎曼几何。球面上没有直线，但有短程线，球面上的短程线就是“大圆周”。我们过球面上的两点和球心，做一个平面，此平面在球面上截出的曲线就是大圆周。例如，赤道和地球上所有的经线都是大圆周，但除赤道以外的所有纬线都不是大圆周，不是短程线。

例如，我们坐飞机由北京直飞纽约的航线，从北京起飞后飞机不是向正东飞，而是向东北方向飞，经过我国的东北，俄罗斯的远东地区，飞到白令海峡附近，然后沿阿拉斯加的北海岸向东，再转向东南，穿越加拿大的大片地区，最后飞抵纽约。

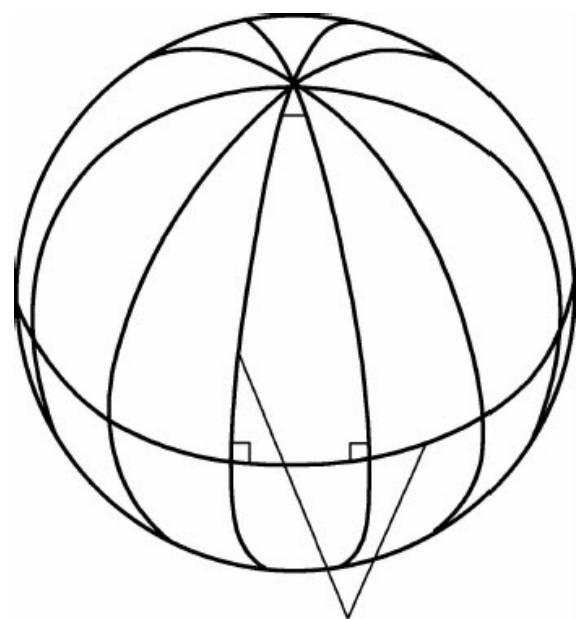
有人或许会问：“干嘛要绕大圈呢？从北京直接往东飞不是近一点吗？”不对，因为从北京直接向东飞往纽约的航线不是最短的，不是短程线。短程线是过北京、纽约和地心三点的平面，截地球表面得到的线，那正是现在北京直飞纽约的航线。所以飞机恰好是沿短程线飞行的。

平面几何中“过直线外的一点能引一条平行线”，在球面情况就是过大圆周外一点，能否再画一个大圆周和自己不相交？答案是“不能”。你想，过赤道外一点，能画出另一大圆周并使它与赤道不相交吗？肯定不能！所以球面上不存在“平行线”。

平面上的三角形是由三条直线围成的，球面上的三角形也应该由三条短程线（大圆周）围成。大家看图3-6，上面标出了赤道和两条经线围成的三角形。由于所有的经线都与赤道垂直，所以此三角形二底角之

和已是180°，再加上顶角，三角形三内角之和肯定大于180°了。

下表列出了三种几何的特点。



大圆周（短程线）
图3-6 球面上的短程线和三角形

	空间曲率	平行线	三角形三内角之和	圆周率	例
欧氏几何	零	一条	$=180^\circ$	$=\pi$	平面
黎氏几何	正	无	$>180^\circ$	$<\pi$	球面
罗氏几何	负	两条以上	$<180^\circ$	$>\pi$	伪球面、马鞍面

1845年，黎曼从更高的角度把这三种几何统一起来，成为黎曼几何，用来描述弯曲和扭曲的几何客体。黎曼曾用这一工作，在哥廷根大学做了申请一个讲师职位的求职报告。

黎曼天才地预见到，真实的空间不一定是平直的。如果不平直，就不能用欧氏几何来描述，而要用黎曼几何来描述。他还预见，物质的存在可能造成空间的弯曲。

黎曼几何为爱因斯坦建立他的广义相对论，准备了数学基础。

探究弯曲的时空

这时，爱因斯坦产生了与当年黎曼类似的猜想。而且，此时的爱因斯坦已经掌握了大量的物理知识，创建新理论的条件已经成熟，这些都是当年黎曼不可能具备的。

起初爱因斯坦与格罗斯曼合作，学习和掌握黎曼几何，并寻找联系物质和时空几何的基本方程——场方程。他们尝试写出了一些形式的方程，但都有重大缺陷。爱因斯坦到德国后，又与希尔伯特探讨。希尔伯特不愧是一位数学大师，爱因斯坦与他作了短时间的探讨，几个月后爱因斯坦就给出了场方程（广义相对论的核心方程）的正确形式，建立起他的新理论——广义相对论。

新理论克服了旧理论的两个基本困难，用广义相对性原理代替了狭义相对性原理，并且包容了万有引力。爱因斯坦认为，新理论是原有相对论的推广，因此称其为广义相对论，而把原有的相对论称为狭义相对论。

应该指出，希尔伯特在与爱因斯坦讨论后不久，也几乎同时得到了类似的场方程。但希尔伯特只是在数学形式上得到了这个方程，并不了解它的深刻物理内容，而且，他对所得到的场方程的物理解释并不完全正确。

爱因斯坦与希尔伯特在1915年底形成了竞争。爱因斯坦关于广义相对论的论文是11月25日完成并投稿的，于12月5日发表，文中给出了正确的场方程。希尔伯特的有关论文是11月20日完成并投稿，1916年3月1日刊出的。希尔伯特的论文投稿时间比爱因斯坦早5天，但稿中没有给

出正确的场方程。他在修改清样期间，看到了爱因斯坦的论文，就在自己的论文中补入了正确的广义相对论场方程。此外，希尔伯特投稿前曾有一封给爱因斯坦的信，祝贺他算出了水星轨道近日点进动的正确值。

可见，爱因斯坦得到广义相对论场方程比希尔伯特要早。不过场方程的数学形式十分复杂，如果没有希尔伯特的帮助，恐怕爱因斯坦很难在1915年底找到场方程的正确形式，并完成这一理论。

希尔伯特曾在给爱因斯坦的信中谈到“我们的理论”，爱因斯坦对此很不高兴，回信说“这是我的理论，什么时候成了‘我们的理论’了？……”希尔伯特后来也承认爱因斯坦是广义相对论的唯一创建人。这个不愉快的小插曲没有影响两人后来的友谊。

走向广义相对论

实际上广义相对论的建立是一个漫长的过程。最初，爱因斯坦企图把万有引力纳入狭义相对论的框架，几经失败使他认识到此路不通，反复思考后，他产生了等效原理的思想。

爱因斯坦曾回忆这一思想产生的关键时刻：“有一天，突破口突然找到了。当时我正坐在伯尔尼专利局办公室里，脑子忽然闪现了一个念头，如果一个人正在自由下落，他绝不会感到自己有重量。我吃了一惊，这个简单的思想实验给我的印象太深了。它把我引向了引力理论。……”

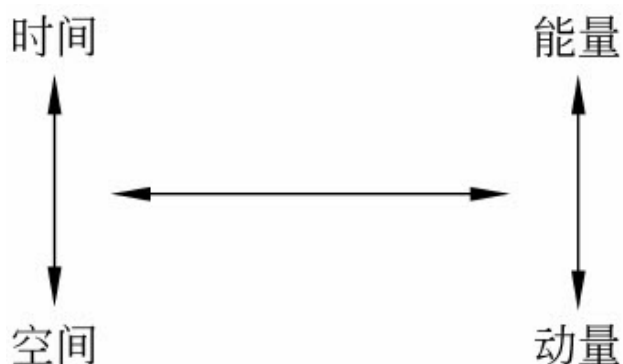
爱因斯坦1905年开始研究万有引力，1907年提出等效原理，1911年得到光线在引力场中弯曲的结论，1913年与格罗斯曼一起把黎曼几何引进引力研究，1915年与希尔伯特讨论，并在当年找到了场方程的正确形式。除去在数学上曾得到希尔伯特和格罗斯曼的有限、然而十分可贵的帮助之外，爱因斯坦几乎单枪匹马奋斗了10年，才把广义相对论的框架大体建立起来。

1905年发表狭义相对论时，有关的条件已经成熟，洛伦兹、庞加莱等一些人，都已接近狭义相对论的发现。

而1915年发表广义相对论时，爱因斯坦则远远超前于那个时代所有的科学家，除他之外，没有任何人接近广义相对论的发现。所以爱因斯坦自豪地说：“狭义相对论如果我不发现，5年之内肯定会有人发现；广义相对论如果我不发现，50年之内也不会有人发现。”

物质告诉时空如何弯曲

广义相对论，实际上是一个关于时间、空间和引力的理论。狭义相对论认为时间、空间是一个整体（四维时空），能量、动量是一个整体（四维动量），但没有指出时间-空间与能量-动量之间的关系。广义相对论进一步指出了这一关系，认为能量-动量的存在（也就是物质的存在），会使四维时空发生弯曲！万有引力并不是真正的力，而是时空弯曲的表现！如果物质消失，时空就回到平直状态。



爱因斯坦给出了广义相对论的基本方程，这个方程被称为爱因斯坦场方程，

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu} \quad (3.7)$$

式中常数 κ 与万有引力常数 G 有关

$$\kappa = \frac{8\pi G}{c^4}$$

其中 c 是光速。爱因斯坦场方程是张量方程，式中带有下角标字母的 $R_{\mu\nu}$ 、 $T_{\mu\nu}$ 、 $g_{\mu\nu}$ 以及 R 都是张量。

爱因斯坦之所以采用张量来表述广义相对论，是因为张量方程在坐标变换下形式不变，他认为这符合自己的广义相对性原理：物理规律不依赖于坐标系的选择。我们不想在此处做过于专门的讨论，感兴趣的读者可参看任何一本介绍广义相对论的书籍。

用爱因斯坦场方程，可以精确地算出，能量-动量的存在，如何影响时空的弯曲。该方程左端是描述时空曲率的量，右端是描述能量-动量的量：

$$\text{时空曲率} = \text{能量动量}$$

实际上，这是由10个二阶非线性偏微分方程组成的方程组，非常难解。



绘画: 张京

时空告诉物质如何运动

相对论把四维时空中的曲线称为世界线。广义相对论认为，万有引力不是一般的力，而是时空弯曲（图3-7）造成的几何效应。质点在万有引力作用下的运动（例如地球上的自由落体；行星的绕日运动等），没有受到力，是弯曲时空中的自由运动——惯性运动。它们在时空中描出的世界线，虽然不是直线，却是直线在弯曲时空中的推广——“测地线”。粗略地说，测地线就是短程线，即两点之间的最短线或最长线（注意，相对论中把最短线和最长线都称为短程线）。当时空恢复平直时，测地线就成为通常的直线。

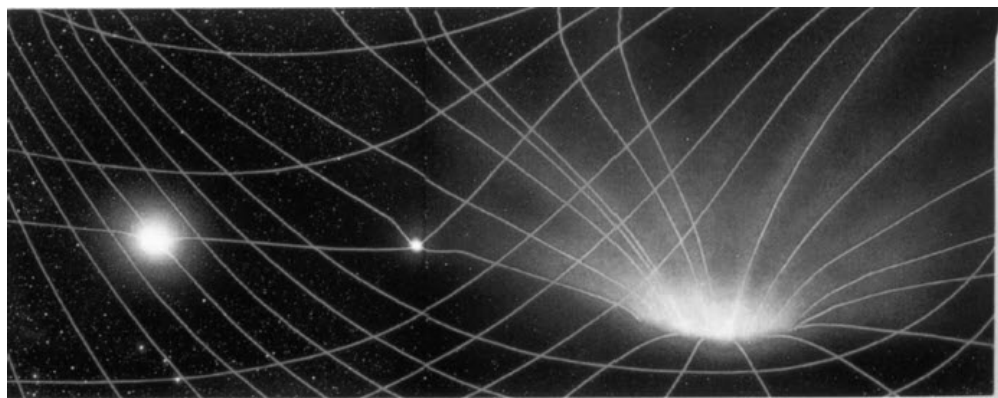


图3-7 弯曲的时空示意图

需要说明的是，在通常的平直空间或正曲率空间中，两点之间存在最短线。但在相对论的四维时空中，两点之间有的有最短线，有的却没有最短线，只有最长线。例如自由质点描出的测地线，实际上是两点间最长的世界线，而不是最短线。

双生子佯谬问题中留在地球上的人近似处于惯性运动状态，它描出的世界线就是测地线，是两点间的最长线。也就是说，在两个确定的时空点之间运动的一群人，其中处于惯性运动状态的那个人描出的世界线

最长。由于世界线的长度就是他经历的真实时间，所以他经历的时间也就最长。星际旅行者经历加速过程，他的世界线不是测地线，所以他会比留在地球上的同胞兄弟年轻。

爱因斯坦采用数学家们已经得到的测地线方程，作为决定弯曲时空中自由质点如何运动的“运动方程”

$$\frac{d^2 x^\alpha}{ds^2} + \Gamma_{\mu\nu}^\alpha \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0 \quad (3.8)$$

方程中s是测地线的弧长， $\Gamma_{\mu\nu}^\alpha$ 称为“联络”，-描述引力场强或惯性场强。

场方程表示“物质告诉时空如何弯曲”，运动方程则表示“时空告诉物质如何运动”。

爱因斯坦初建广义相对论时，认为广义相对论的基本方程有两个：场方程（3.7）和运动方程（3.8）。后来，爱因斯坦和苏联的福克分别证明，从场方程可以推出运动方程，因此，广义相对论的基本方程只有一个——场方程（3.7）。

另外，在他们的证明中还得到一个值得注意的副产品：场方程中作为场源的质量，在推出的运动方程中，同时出现在惯性质量和引力质量两个位置上。这告诉我们，在广义相对论的理论框架中，引力质量和惯性质量是同一个东西。

如何理解时空弯曲

我们打个比方来较为形象地说明时空弯曲。假如四个人各拉紧床单的一个角，床单这个二维空间就是平的。放一个小玻璃球在上面，如果不去推它，它就会保持静止或匀速直线运动状态不变（假设床单足够光滑，床单的微小摩擦力可以忽略）。如果床单中央放一个铅球，床单就会凹下去，这个二维空间就弯曲了。这时，如果再放置一个小玻璃球在床单上，它就会滚向中央的大球。

在这个例子中，可以把大球看做“地球”，小球看做一个下落的物体。小球为什么会滚向大球呢？按照牛顿的观点，这是由于大球（地球）用“万有引力”吸引小球。按照爱因斯坦的观点，则是由于大球（地球）的存在使空间弯了，并不存在什么“引力”，小球落向大球乃是弯曲空间中的自由（惯性）运动。

如果给小球一个横向速度，它就会绕大球转起来。这时可把大球看作太阳，小球比作行星。为什么小球（行星）不远离大球（太阳）飞向远方呢？按照牛顿的观点，这是由于小球受到大球的“引力”，不能跑向远方，只能环绕大球运动。按照爱因斯坦的观点，小球并未受到任何力，只是由于空间弯曲了，在弯曲空间中它作自由（惯性）运动不能飞向远方。

对上述比喻应该加以解释的是，上面例子所说的只是“空间”弯曲，而广义相对论说的则是四维“时空”的弯曲。太阳的存在使四维时空弯曲了，行星绕日运动，就是在弯曲时空中的惯性运动，行星描出的世界线是四维时空中的测地线，根本就不存在什么万有引力。

注意，这里所说的测地线不是指行星在三维空间中的椭圆轨道，而是指图3-8中所示的行星在四维时空中描出的螺旋状世界线。图3-8是在相对于太阳静止的参考系中绘出的时空图。太阳描出的世界线是一条与时间轴平行的直线，而行星绕日运动的世界线则是一条螺旋线。

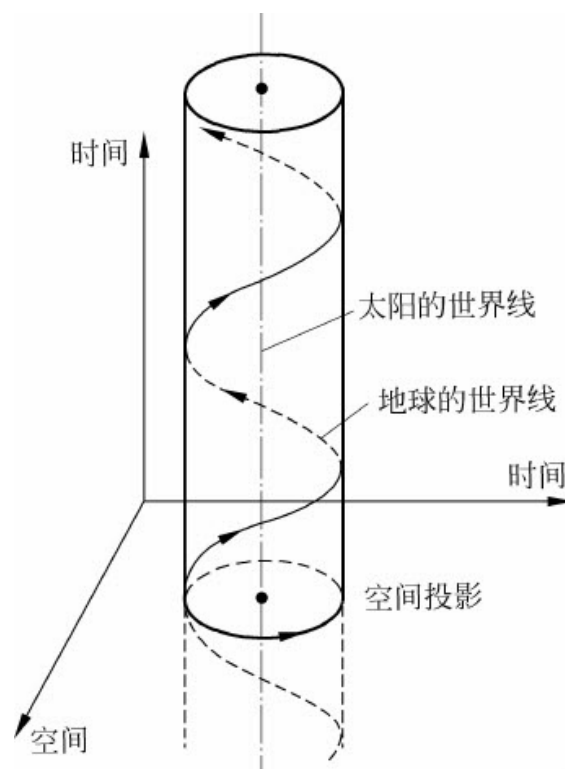


图3-8 四维时空中行星绕日的运动

广义相对论的实验验证

爱因斯坦发表广义相对论的时候，求出了场方程的一些近似解。他在发表自己理论的时候，同时提出了三个检验广义相对论的实验：

（1）引力红移；（2）行星轨道近日点的进动；（3）光线偏折。这3个实验均被观测证实。除去这3个验证实验外，还有1970年前后所做的雷达回波延缓实验，以及1978年发现的脉冲双星运转周期减小，从而间接证实引力波存在的实验。

引力红移

按照广义相对论，时空弯曲的地方，钟走得慢，即时间会变慢。时空弯曲得越厉害，钟走得越慢，所以，太阳附近的钟，会比地球上的钟走得慢。但是我们不可能在太阳表面放一个钟，即使放一个钟也不敢用望远镜去看，太阳光实在太强了。不过没有关系，太阳表面原本就有钟。

我们知道，每种元素都有特定的光谱线。一根频率为 ν 的光谱线，表示原子内部有一个以频率 ν 走动的钟。太阳表面有大量氢原子，因此可以比较太阳附近氢原子发射的光谱线和地球实验室中的氢光谱线来进行检验。由于太阳附近的钟变慢，那里射过来的氢原子光谱线（与地球上的氢光谱比较）频率会减小，即谱线会向红端移动。这就是广义相对论预言的引力红移，它反映太阳表面的钟变慢。后来的观测实验证实了这一预言。

此后，人们又检验了银河系中其他一些恒星的引力红移。白矮星等高密度恒星的引力红移远比太阳强，但这些恒星离我们太远，观测精度较低。后来，人们又利用穆斯堡尔效应在地面上检验了引力红移。上述检验的结果均与广义相对论的预言一致。

近年来的研究表明，宇宙学红移（即天文观测发现的河外星系远离我们造成的红移现象）不属于多普勒效应，也属于引力红移。

行星轨道近日点的进动

爱因斯坦谈到的第二个检验广义相对论的实验是：行星轨道近日点的进动。

牛顿的万有引力定律算出，行星的轨道是一个封闭的椭圆，正好与开普勒定律相符。然而，实际观测表明，行星轨道不是一个封闭的椭圆，轨道的近日点不断向前移动（进动），如图3-9所示。这个效应以离太阳最近的水星最为显著，每百年高达5600弧秒，这种效应主要可归因于天文学上的岁差，以及其他行星对水星运动的影响。扣除这些影响后，尚有约43弧秒/百年的进动无法解释。

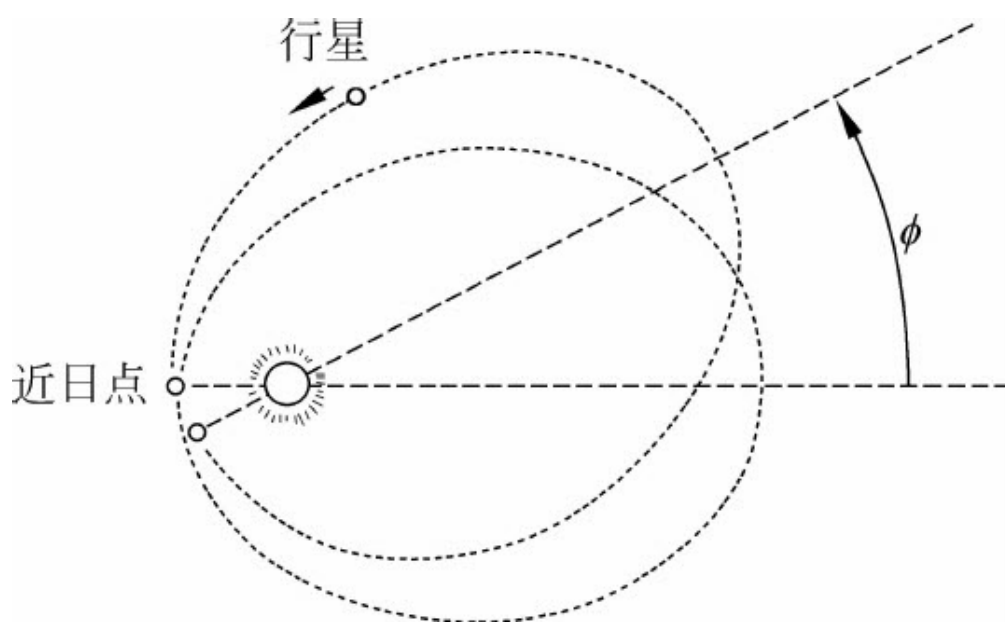


图3-9 行星轨道进动

按照广义相对论，行星绕日的轨道不再是封闭的椭圆，轨道的近日点会不断向前移动。

当时许多人怀疑存在一颗比水星离太阳更近的未知行星，而水星轨道的剩余进动就来源于这颗星的影响。法国天文学家勒维叶在预言海王

星成功之后，曾通过水星轨道的这一偏差来反推出这颗离太阳极近的未知行星的轨道。曾有一度，人们把太阳盘面上移动的一个黑点误认为是这颗未知的行星，由于它离太阳这个火球非常近，给它起名为火神星。然而不久就发现那不过是太阳表面的一个黑子，所谓的火神星纯属子虚乌有。此后，水星轨道近日点每百年43弧秒的进动一直是个未解之谜。

广义相对论算出的行星轨道，本身就不是一个封闭的椭圆，不需其他行星影响，自己就会“进动”。而且，对于水星轨道，这个进动值恰恰就是每百年43弧秒。这样，实验观测支持了广义相对论。而且，这一实验是所有验证广义相对论的实验中精度最高的。

爱因斯坦在完成广义相对论之前，就知道水星轨道这43弧秒的进动值一直没有得到解释。当他用广义相对论算出这一进动值时，高兴极了，他在给洛伦兹的信中说：“我现在正为历尽艰辛获得的理论的清晰，以及它与水星轨道近日点进动的一致而感到快乐。”他在给其他友人的信中说：“（在发现自己的理论与水星轨道进动值密切相符后）我简直高兴极了，一连几个星期我都高兴得不知怎么样才好。……”这毫不奇怪，因为物理学（其他自然科学也一样）是一门实验与测量的科学，只有能够被实验和测量严格定量证实的理论才能进入物理学。

爱因斯坦认识到，广义相对论与水星轨道运动的这种高度一致性，表明自己已经找到了正确的引力场方程，自己的理论已经成功了。

太阳附近的光线偏折

爱因斯坦检验广义相对论的第三个观测实验是光线偏折。由于太阳造成时空弯曲，遥远恒星的光通过太阳附近时会发生偏折，弯向太阳（图3-10）。虽然从牛顿的万有引力定律也可得出光线偏折的结论，但其偏转角只有广义相对论预言值的一半。

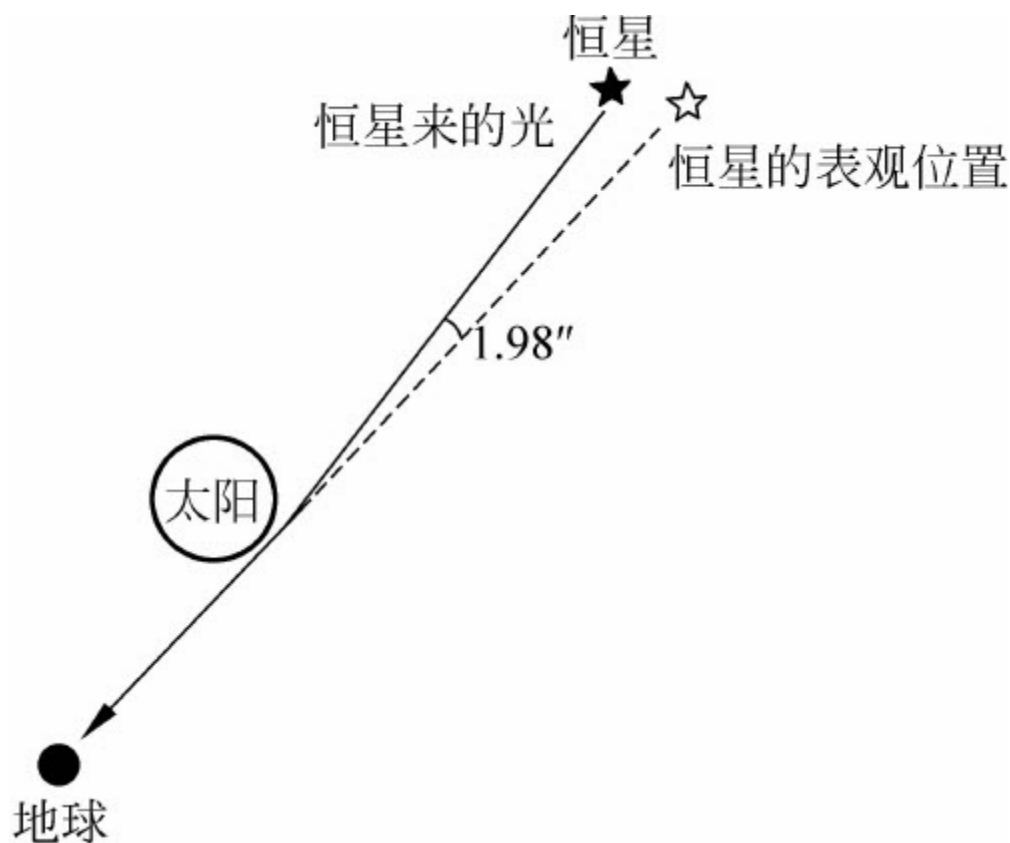


图3-10 光线偏折

这一观测很难进行，需要拍下太阳背后的星空，来与太阳不存在时的同一星空照片比较，观察并测量恒星位置的偏离。太阳比恒星亮得多，白天根本不可能拍下太阳背后的星空。唯一的可能是在发生日全食的时候进行拍摄。当月亮的影子完全挡住太阳，太阳背后的恒星在黑暗中显现的时候，抓紧拍下照片。不存在太阳的同一星空背景，则需在几

个月前或几个月后拍摄。

平常我们看到太阳每天从东方升起到西方下落一次，这叫太阳的周日运动（地球自转引起）。此外，太阳还有一个周年运动（地球公转引起），即每天的同一时刻，太阳在星空背景上的位置都不同，都要移动差不多一度，全年正好移动一圈。所以，白天出现在太阳背后的星空，几个月前或几个月后，会在夜间出现。

英国的爱丁顿在1919年日全食的时候，首次进行了检验光线偏折的观测。两支观测队分别到达将出现日全食的不同地点，南美洲的巴西和非洲西岸的普林西比。

爱丁顿亲自率领的一队，在普林西比碰上阴天，幸运的是在日全食即将结束之前，一阵风吹开了乌云。他们在6~8分钟的日全食时间内，拍了15张照片。

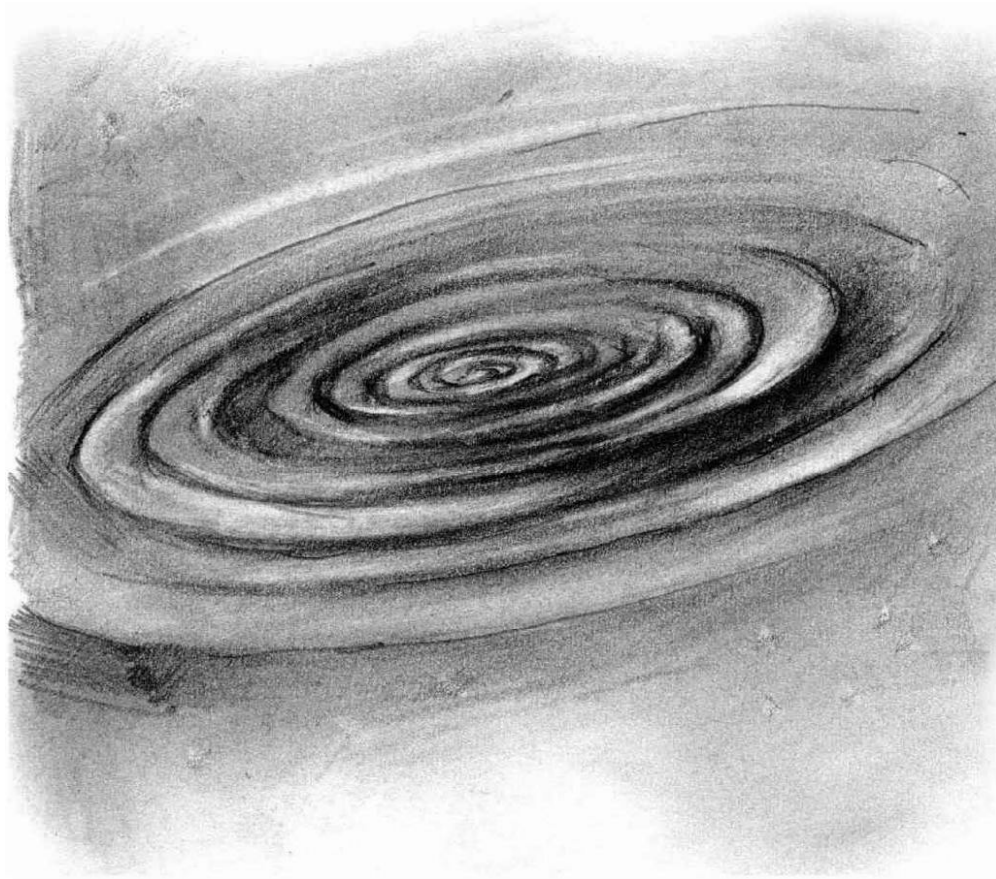
去巴西的那一队遇到艳阳天，比较顺利地拍摄了照片。但后来发现，由于仪器过热造成了胶片形变，幸亏经过修正后，数据依然可用。几个月后太阳移开了这一星空区，他们又拍了这一星空区的照片。

从照片上比较，光线确实偏折了，两队观测数据显示的偏转角分别为1.61秒和1.98秒，接近广义相对论预言的1.75秒，而比牛顿万有引力定律预言的0.88秒大一倍左右。观测支持了广义相对论。

消息传到德国时，有人问爱因斯坦有什么感想，他平静而自信地说：“我从来没有想过会是别的结果。”光线偏折实验是对广义相对论的有力支持。

严格而美妙的数学物理体系，高深难懂的黎曼几何和张量分析，精

密神奇的实验验证，再加上爱因斯坦发表狭义相对论和光子说的巨大影响，使广义相对论一下就得到了科学界的承认，爱因斯坦的威望也达到了一生中的顶峰。



绘画: 张京

第四章 奇妙的黑洞

我们现在从广义相对论的角度来描述黑洞，也就是说，介绍一下弯曲时空中的黑洞。这需要一点数学知识。事实上用一点数学，比完全不用数学，能使普通读者更容易、更清楚也更正确地了解黑洞。具备一点最初步的高等数学知识的读者，就可以读懂下面的讨论。

四维时空

我们先介绍一下四维时空的概念。爱因斯坦的狭义相对论发表之后，他大学时代的数学老师闵可夫斯基发现，如果把三维空间和一维时间看成一个整体的四维时空，就可以把爱因斯坦的狭义相对论写成更清楚、更简洁的形式。

大家知道，在笛卡儿的直角坐标下，三维空间中两点之间的距离 dl 可以写成

$$dl^2 = dx^2 + dy^2 + dz^2 \quad (4.1)$$

如果换成球坐标表述，则可写成

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2 \quad (4.2)$$

闵可夫斯基把时间看做第四维空间，把四维时空中两点之间的距离（在相对论中称为间隔） ds 写成

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \quad (4.3)$$

这是直角坐标系下的表达式，如果换用球坐标，则应写为

$$ds^2 = -c^2 dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2 \quad (4.4)$$

注意，时间项的前面与3个空间项的前面差一个负号，这是“光速不变”原理决定的。在此我们就不详细介绍了，感兴趣的读者可参看介绍相对论的书籍。

式（4.3）与式（4.4）表达的四维时空，称为闵可夫斯基时空。它

不属于欧几里得空间，而属于伪欧几里得空间，关键就在于时间项与空间项前面的符号有正有负。

应该说明，上面4个式子虽然有的是直角坐标有的是球坐标，但都表示的是平直时空，采用球坐标并不表示时空一定弯曲。

闵可夫斯基小时候是个神童，他与数学家希尔伯特是同学。他聪明到什么程度呢？聪明到让小希尔伯特对自己都缺乏信心了，觉得自己大概不行，闵可夫斯基兄弟几人才是真正的聪明人。长大后，希尔伯特成为数学大师，闵可夫斯基的成就却不算大。

这次，闵可夫斯基抓住自己学生爱因斯坦的相对论，作出了一个重要贡献。他创建的四维闵可夫斯基时空，为爱因斯坦后来把自己的相对论，发展成广义相对论铺下了第一块数学基石。爱因斯坦在刚刚看到闵可夫斯基的这一工作时，开玩笑说：你这么一搞，我都看不懂自己的相对论了。

世界线

四维闵可夫斯基时空中的一个点，用 (t, x, y, z) 四个坐标表示，称为一个事件。三维空间中的一个点，不管是运动的点还是不动的点，由于时间的不停发展，在四维时空中都会描出一根线，称为世界线。

图4-1中A、B、C三条世界线，A描述三维空间中的一个不动点，B描述一个匀速直线运动的点，C描述一个变速运动的点。 ds 为世界线上两点之间的“距离”。注意，由于不可能在图上画出时空的全部四个维度，我们没有画出 z 坐标描述的那一维空间。

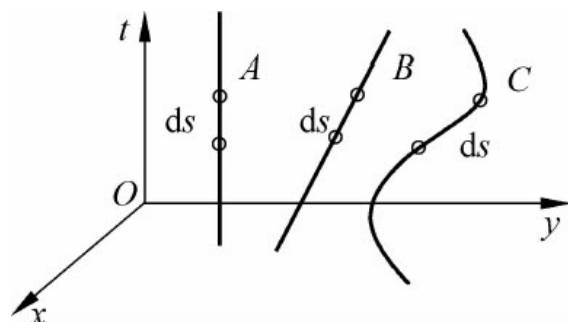


图4-1 四维时空中的世界线

ds 通常称为两点的间隔。由于两点之间总可以用世界线相连，所以 ds 又可看作世界线的线元。从式（4.3）不难看出，当 $ds^2=0$ 时，有

$$v^2 \frac{dx^2 + dy^2 + dz^2}{dt^2} = c^2$$

这表明从点1（事件1）到点2（事件2）的运动速度正好是光速，这段间隔 ds 恰好描述光信号的运动。我们称此间隔为类光间隔。我们称描述亚光速信号（ $v^2 < c^2$ ）的间隔为类时间隔；描述超光速信号（ $v^2 > c^2$ ）的间隔为类空间隔。从式（4.3）不难看出

$$ds^2 > 0 \quad v^2 > c^2 \quad \text{间隔类空}$$

$$ds^2 = 0 \Leftrightarrow v^2 = c^2 \quad \text{间隔类光}$$

$$ds^2 < 0 \quad v^2 < c^2 \quad \text{间隔类时}$$

光锥

时空中任选一点 P ，与 P 点的间隔类光的点组成如图4-2所示的锥面，称为 P 点的光锥。光锥实际上是四维时空中的一个三维超曲面，我们略去了一维空间。

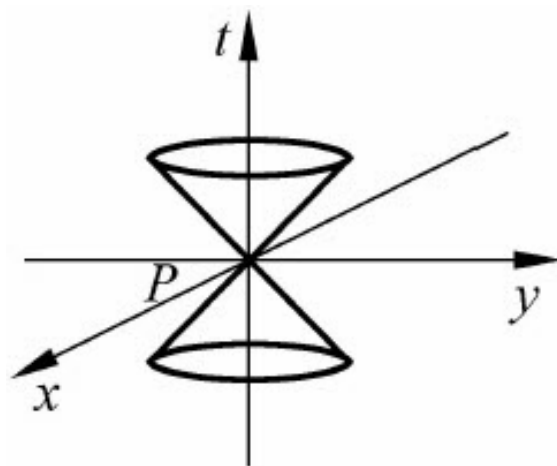


图4-2 光锥图

光锥内部的点与 P 点的间隔都是类时的，与 P 点以亚光速信号相联系。上半光锥内部的点，处在 P 点的未来，从 P 点出发的亚光速信号或质点可以到达它们。下半光锥内部的点处在 P 点的过去，从那里出发的亚光速信号或质点可以抵达 P 点。

上半光锥面上的点也处在 P 点的未来，从 P 点出发的光信号可以抵达它们。下半光锥面上的点处在 P 点的过去，从那里出发的光信号可以抵达 P 点。

光锥外部的点与 P 点的间隔类空，只有超光速信号或质点才能抵达它们，或从它们抵达 P 点。相对论认为光速是信号的最大传播速度，不存在超光速的信号传播和物体运动。所以，光锥外部的点与 P 点没有因果联系。

总之，只有光锥面上和光锥内部的点与 P 点有因果关系。下半光锥（内部及面上）的所有点，都处在 P 点的因果过去，都可以对 P 点的事件产生影响。上半光锥（内部及面上）的所有点，都处在 P 点的因果未来， P 点的事件都可以影响它们。

P点的一个矢量如果落在光锥内部，称其为类时矢量；如果落在光锥外部，称其为类空矢量；如果恰好落在光锥面上，则称其为类光矢量（图4-3）。

一条世界线，如果线上每一点的切矢量都类时，称其为类时世界线，简称类时线。如果每一点的切矢量都类空，称其为类空线；如果每一点的切矢量都类光，称其为类光线。

显然，过P点的类时线都在光锥内部，类光线都在光锥面上，类空线都在光锥面外。

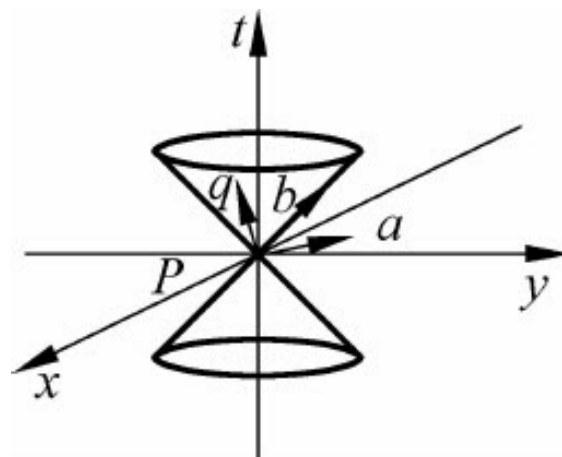


图4-3 P点的类时矢量（ Pq ）、类光矢量（ Pb ）和类空矢量（ Pa ）

球对称的弯曲时空

爱因斯坦发表广义相对论时，自己并没有得到场方程的严格解，他是用近似解算出三大实验验证的结果的。

不过，他1915年底完成广义相对论，第二年，德国数学家兼天文学家史瓦西，就求出了广义相对论场方程的第一个有意义的严格解。即当时空中存在一个不变化的球对称质量，而且球体外面是真空时，外部时空将如何弯曲。他的结论是，这时，时空中两点之间的距离将变成

$$ds^2 = - \left(1 - \frac{2GM}{c^2 r}\right) c^2 dt^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2$$

(4.5)

大家看，与式（4.4）比较，式（4.5）右边的前两项中多了两个带括弧的因子。正是这两个因子的存在，表明时空弯曲了。这就是著名的史瓦西解，它反映了一个不随时间变化的球对称星体外部的时空弯曲情况。

当时第一次世界大战正在紧张进行，应征入伍处在前线的史瓦西得了一种不治之症，被送回后方医院治疗。他在住院期间完成了这一研究。由于他自己行动不便，爱因斯坦代表他在德国皇家学会上宣读了这篇论文。爱因斯坦高度赞扬了他的工作。

遗憾的是，不久之后史瓦西就离开了人世。这场疾病夺去了他年轻的生命，但也为他提供了做出这一成果的机会。否则，战争的枪林弹雨有可能同时剥夺他的生命和科研机会，世界将永远不知道存在过这样一位杰出的学者。

我们要再强调一下，式（4.5）所示的史瓦西解是一个“真空解”，即

球状物体外部全是真空，此解描述的是球体外部真空区的时空弯曲情况。式（4.5）并不反映球状物体内部的时空弯曲情况。

球体内部的时空要用所谓“史瓦西内解”描述。不过，为了抗拒物质间的“万有引力”，必须设想球体内部有某种“排斥力”，许多学者设想了不同的排斥力，结果得到了不同的“史瓦西内解”。

因此史瓦西内解并不唯一，而描述球体外部时空弯曲的“外解”则是唯一的。这就是说，式（4.5）所示的史瓦西解（即外部解）是唯一的。人们最感兴趣的是这个外部解，因为它在天文学和物理学上大有用处。

我们再来分析一下史瓦西解式（4.5）。不难看出，当 $M=0$ 时，式（4.5）回到了平直时空的式（4.4）。这表明，物质不存在时，时空是平直的，这当然是正确的。另外，当 $r \rightarrow \infty$ 时，式（4.5）也回到平直时空的式（4.4），这表示在无穷远处，在无限远离这个球对称质量的地方，弯曲时空恢复到平直时空，这当然也是合理的。

线元与度规

式(4.2)~式(4.5)称为相应时空中的“线元”表达式。公式右端“坐标微分元”前的系数，称为时空的度规。例如式(4.5)右端的 $-\left(1-\frac{2GM}{c^2 r}\right)$ 、 $\left(1-\frac{2GM}{c^2 r}\right)^{-1}$ 、 r^2 和 $r^2\sin^2\theta$ 就是史瓦西时空中度规的分量。线元和度规，反映了这个时空的弯曲情况和几何性质。求解广义相对论中的爱因斯坦方程，就是要算出这一时空中度规的全部分量。

奇点与奇面

数学物理学家在研究式（4.5）时发现，这个解在

$$r = 0 \quad (4.6)$$

处存在一个奇点，在

$$r = \frac{2GM}{c^2} \quad (4.7)$$

处，存在一个奇面（图4-4）。不难看出，当 $r=0$ 时， dt^2 前的系数 $\left(1 - \frac{2GM}{c^2 r}\right)$ 为无穷大；当 $r = \frac{2GM}{c^2}$ 时， dr^2 前的系数虽然不发散， dr^2 前的系数 $\left(1 - \frac{2GM}{c^2 r}\right)^{-1}$ 却发散了。产生无穷大的地方，时空似乎出现了奇异性。

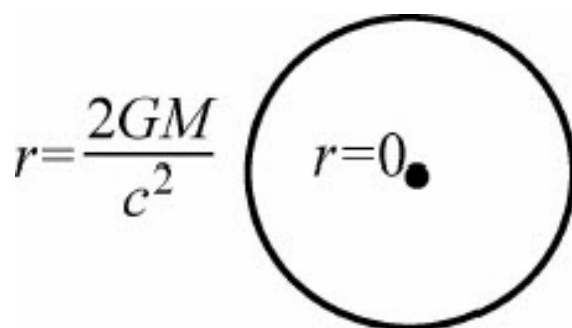


图4-4 史瓦西时空的奇点和奇面

进一步的研究表明， $r=0$ 处的奇异是真奇异，那里时空曲率为无穷大，而且这种无穷大不能通过坐标变换来消除，不管选择什么坐标系， $r=0$ 都是奇点，这种奇点叫内禀奇点。

而 $r = \frac{2GM}{c^2}$ 处的奇面却不一样，那里的时空曲率并不发散，也就是说曲率正常，而且换一个坐标系（例如选择自由下落的坐标系），奇面就会消失。所以奇面处的奇异是假奇异，我们称其为“坐标奇异性”，这种奇异性是由于坐标系选择得不好而造成的，并非时空的固有（内禀）性质。

不过，后来的研究表明，这个奇面虽然是假奇异，但却有重要的物理意义。不难看出，它是球对称黑洞的表面，也就是式（1.4）给出的暗星的表面，它的半径称为引力半径，通常记为 r_g 。

时空坐标互换

我们来介绍黑洞的一个奇怪性质：在黑洞内外，时空坐标会发生互换。在黑洞内部原来的时间坐标 t 会变成空间坐标，原来的空间坐标 r 则会变成时间坐标。

我们先看一下式（4.5），它与式（4.4）有相似之处，时间项 dt^2 前的系数是负的，而空间项 dr^2 、 $d\theta^2$ 和 $d\phi^2$ 前面的系数都是正的，这正是四维时空中时间与空间的区别。然而，上述结论仅在

$$r > \frac{2GM}{c^2} \quad (4.8)$$

的黑洞外部是正确的。在黑洞内部，由于，

$$r < \frac{2GM}{c^2} \quad (4.9)$$

括号中的因子变成了负的，这导致 dt^2 前的系数变成了正的，与 $d\theta^2$ 、 $d\phi^2$ 等空间项前的系数符号相同了，而 dr^2 前的系数却变成了负的。这表明，在黑洞内部， r 变成了时间坐标， t 变成了空间坐标，于是造成了黑洞内部时空坐标的互换。

单向膜区与表观视界

由于洞内的 r 变成了时间，洞内的等 r 面不再是空间的球面，而变成了时间的“等时面”。时间是流逝的，有方向的。任何物质都必须顺着时间的方向前进，都必须“与时俱进”。

如果时间的方向向里，洞内的任何物质都不能停留，都必须向 $r=0$ 处会聚。不过这时 $r=0$ 已不再是球心，而变成了时间的“终点”。所有的 $r=$ 常数的等时面实际上都成了只能向里透的单向膜。物质到达时间的终点 $r=0$ ，在那里会处于时间之外。处于“时间之外”是什么意思，现在还不清楚。

如果洞内的时间 r 方向向外， $r=0$ 就是时间的起点， r 为常数的等时面成为了向外的单向膜。物质在时间的起点 $r=0$ 处产生（无中生有），然后通过单向膜向外运动，最终抛出洞外。

研究认为，时间向里的洞，就是黑洞，时间向外的洞则是白洞。广义相对论的解只表示 $r=\frac{2GM}{c^2}$ 以里是“洞”，但没有限定它一定是黑洞，还是白洞。

不过，一般认为，这种“洞”是恒星塌缩形成的，形成“洞”的初始情况是物质向内塌陷，这样形成的洞应该是黑洞。所以一般人只讨论黑洞，很少有人讨论白洞，但要指出，广义相对论并不排斥白洞的存在。

黑洞是一切物质和辐射都向里掉，任何物质和辐射都跑不出来的星体，白洞则是不断往外喷东西，但任何东西都掉不进去的星体。

黑洞和白洞的内部都是单向膜区，所以洞内都是真空，任何物质都

不能在单向膜区停留。洞的表面 $r=\frac{2GM}{c^2}$ 是单向膜区的起点。单向膜区的起点称为表观视界。

黑洞内外的光锥图

我们现在画出黑洞内外的光锥图（图4-5），它能使我们更好地理解单向膜区。研究表明，由于时空弯曲，黑洞外部的光锥，越靠近黑洞表面，光锥越扁。在黑洞表面上，光锥退化为一根线。到黑洞内部，由于时空坐标互换， r 变成了时间，光锥横过来了，随着靠近奇点 $r=0$ ，光锥由“胖”变“瘦”。在“洞”外，光锥的未来指向是向上的，这是由时间 t 的正向决定的。

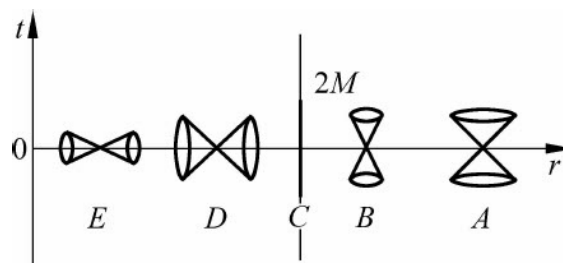


图4-5 视界内、外的光锥图

但在“洞”内，我们无法简单地用时间 r 来判断光锥的未来指向。实际上，这要由“洞”形成的初始条件来决定。因为黑洞是塌缩形成的，物质落向星体的中心，这一初始条件决定黑洞内部的未来光锥指向 $r=0$ 的方向，也即时间 r 指向 $r=0$ 的方向。我们把未来光锥指向 $r=0$ 的“洞”称为黑洞，未来光锥背向 $r=0$ 的“洞”，即时间方向背向 $r=0$ 的“洞”，称为白洞。

无限红移面

前面曾经谈到，广义相对论预言弯曲时空中的钟会变慢，从而导致那里的光源发出的光会出现红移，这种红移称为引力红移（即时空弯曲造成的红移），实验观测支持了这一结论。

为了探讨黑洞的性质，我们来研究一下引力红移效应。广义相对论认为，一个球对称星体（例如太阳）造成的弯曲时空中，时钟变慢由下式决定：

$$\Delta t = \frac{\Delta \tau}{\sqrt{1 - \frac{2GM}{c^2 r}}} \quad (4.10)$$

式中， M 为星体质量， G 为万有引力常数， c 为真空中的光速。 τ 为静止在星体（太阳）附近的弯曲时空中的钟指示的时间， r 为该钟到星体中心的距离； t 为无穷远处（那里时空平直，相对于太阳，地球就可看做是无穷远）观测者的钟所指示的时间。由于公式根号中的因子小于1，所以 $dt > d\tau$ 。这表明太阳表面的钟走1秒时间，地球处的钟走的时间 dt 将多于1秒，因此在地球上的观测者看来，太阳表面的钟变慢了。相应的引力红移公式为

$$\nu = \nu_0 \sqrt{1 - \frac{2GM}{c^2 r}} \quad (4.11)$$

式中， ν 为地球观测者拍到的太阳光谱线的频率， ν_0 则为地球观测者在地球实验室中拍到的同一种元素的同根光谱线的频率。从上式看，显然 $\nu < \nu_0$ ，所以在地球上的人看来，从太阳来的光线的频率减小了，即波长增大了，发生了红移。

我们在第三章中已经谈到，实验观测支持了上述结论。现在我们来
看，当星体不是太阳，而是黑洞，会发生什么情况。

我们从黑洞表面直到观测者所在的位置，放置一系列钟和光源。对于靠近黑洞表面的钟和光源，由于

$$r - r_g = \frac{2GM}{c^2} \quad (4.12)$$

式（4.10）与式（4.11）中根号内的因子趋于零。于是我们看到，位于黑洞表面附近的钟，即使 $d\tau$ 取很小的值，都会有 $dt \rightarrow \infty$ 。所以，在远方观测者看来，黑洞表面处的钟完全不走了。我们从式（4.11）则看到，不管 v_0 取什么值，都会有 $v \rightarrow 0$ ，即波长 $\lambda \rightarrow \infty$ ，光谱线发生无限红移。因此，我们称黑洞的表面为无限红移面。

事件视界

我们看到史瓦西黑洞的表面，既是单向膜区的起点（表观视界），又是无限红移面。它还有一个重要性质：它是零超曲面（类光超曲面），而且是事件视界。

超曲面是四维时空中的三维曲面。零超曲面有一个“怪异”的特点：它虽然有法矢量，但法矢量的长度为零。这似乎不可思议，我们在三维欧氏几何中熟知，任何曲面都有法矢量和切矢量，切矢量“躺在”曲面上，法矢量则与曲面垂直，二者的长度都不可能是零。这次怎么出现了长度为零的法矢量呢？

这种曲面又称为“迷向曲面”，是伪欧空间和伪黎曼空间特有的。大家都知道欧式空间是平直空间，黎曼空间是弯曲空间，欧氏空间可看做黎曼空间在曲率为零时的特殊情况。所谓“伪”，是指在空间任意两点距离的表达式中，各坐标对应项的前面正负号有差别，例如式（4.4）与式（4.5）的时间项和空间项，差了个正负号。

在这类“伪”空间中，有一类曲面，它的法矢量倒在曲面上，与其中一个切矢量重合，这时法矢量的长度就是零。实际上，这种法矢量是类光矢量。在伪黎曼时空中，矢量的长度可以为正，还可以为零为负。长度为零的矢量就是类空矢量，长度为负的矢量是类时矢量，长度为零的则是类光矢量，也即零矢量。

法矢量长度为零的超曲面称为类光超曲面或零超曲面。黑洞的表面就是类光超曲面，称为事件视界。不可能有任何物质或信息从事件视界以内（即黑洞内部）跑出，并到达无穷远。也就是说，远方的观测者得

不到黑洞内部的任何信息。黑洞的表面，是外部观测者能收到信息的“边界”，所以称其为事件视界。

对零曲面的诠释

为了使读者对零曲面有更直观的了解，我们介绍一下狭义相对论中的时空图。

我们在S系的二维时空图中画出S'系的x'轴与t'轴的位置。x'轴是由t'=0决定的，而t'轴又是由x'=0决定的。从洛伦兹变换

$$\begin{cases} x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}} \\ y' = y \\ z' = z \\ t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - v^2/c^2}} \end{cases} \quad (4.13)$$

可知，t'=0导致

$$t = \frac{v}{c^2}x \quad (4.14)$$

而x'=0导致

$$t = \frac{x}{v} \quad (4.15)$$

在S系的t-x图4-6中，式（4.14）描出的就是x'轴，式（4.15）描出的则是t'轴。上二式表明，x'轴与t'轴在t-x图中都是直线，而且斜率都是正的。如果采用c=1的自然单位制，它们的斜率恰好互为倒数。我们在图4-

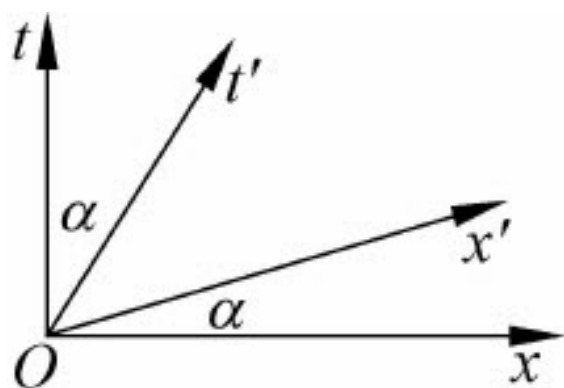


图4-6 在S系中画出的洛伦兹变换的时空

6中画出了上述坐标轴。

图

图4-6给出的是洛伦兹变换下两个坐标系之间的关系，是在S系中画出的，S'系以速度 v 相对S系沿 x 轴正向运动。

显然， t' 轴与 x' 轴是正交的，但从图4-6看，它们似乎不正交。这是闵可夫斯基时空给人的错觉，是由于闵氏时空不是欧几里得时空而是伪欧时空造成的。

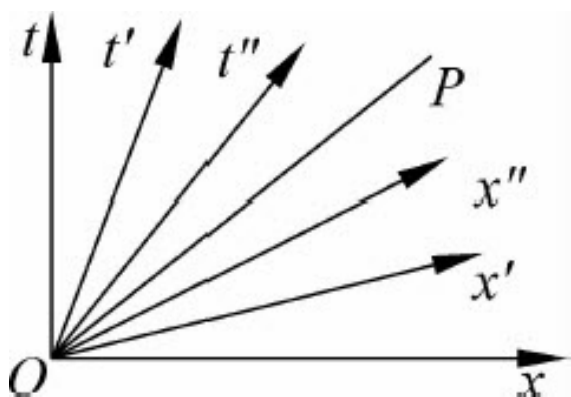


图4-7 闵可夫斯基时空中的正交性

如果有第三个惯性系 S'' ，以更快的速度 u 相对于S系运动，则它的 t'' 轴与 x'' 轴如图4-7所示。 u 越大， t'' 轴与 x'' 轴越接近，当 u 趋于光速 c 时， t'' 轴与 x'' 轴重合在光锥面 OP 上。

我们知道， t 轴可以看作“同时超曲面”（由 $t=\text{常数}$ 决定的四维时空中的三维超曲面，简称同时面）的法矢量，而 x 轴为此超曲面的一个切矢量。同样， t' 轴是同时面（ $t'=\text{常数}$ ）的法矢量， x' 轴为它的一个切矢量； t'' 轴为同时面（ $t''=\text{常数}$ ）的法矢量， x'' 轴为其一个切矢量。

当 $u \rightarrow c$ 时， t'' 轴与 x'' 轴重合于光锥面上。这种法矢量倒在超曲面内，并与其一个切矢量重合的现象在黑洞研究中极为重要。我们称这样的超曲面为零超曲面，或类光超曲面，简称零曲面或类光曲面，它的法矢量称为零矢量或类光矢量。

存在零曲面是闵氏时空的一个特点，与这种时空的伪欧性质有关。通常的欧氏空间不可能存在这种曲面。

我们看到，光锥面是零曲面。在时空中运动的光波的波前是零曲面，黑洞的表面也是零曲面。黑洞所在的时空，虽然不是伪欧的，但是伪黎曼的。“伪”空间都有可能存在零曲面。

不过应该说明，虽然事件视界都是类光超曲面，但类光超曲面不都是事件视界。例如光波的波前，也是类光超曲面，但它不是事件视界。

我们看到，球对称的史瓦西黑洞的表面，既是事件视界，又是表观视界，还是无限红移面。

飞向黑洞的飞船

下面我们就来看一下，如果一艘飞船驶向黑洞，远方的观测者能看到怎样的景象，飞船上的宇航员又能看到或感受到什么（图4-8）。

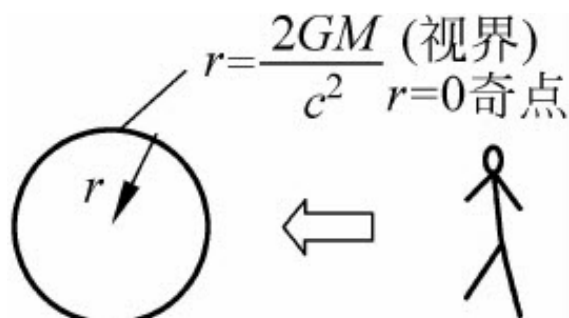


图4-8 驶向黑洞的飞船

如果从观测者到黑洞，在宇宙空间摆放一系列钟和一系列光源。由于时空弯曲的地方钟走得慢，时空弯曲得越厉害，钟就走得越慢。如前面所述，置于黑洞表面的钟将完全停滞不走。而且，时空弯曲得越厉害的地方，光源发出的光红移越大，置于黑洞表面的光源发射的光将发生无限红移，波长 λ 趋于无穷大。

所以，飞船飞向黑洞时，远方的观测者将看到飞船越飞越慢，飞船上人的动作越来越慢，飞船的颜色也越来越红。最后飞船就粘在黑洞的表面上，宇航员的动作也完全僵化，就像雕塑剧一样。但看不见飞船进入黑洞。

飞船进去没有呢？飞船上宇航员用的钟不是摆在飞船外的钟，宇航员的时间进程也不是飞船外的观测者经历的时间进程，宇航员觉得自己在飞船上的钟并没有变慢，飞船很正常地进入了黑洞。

那么远方观测者为什么没有看见飞船进入黑洞，只看见飞船粘在黑洞表面呢？那是因为飞船的“背影”留在了洞外，远方观测者看到的正是这个“背影”。

在地球上，我们看到一个人从屋里走出去，他的背影一闪就消失

了。这是因为组成他背影的光子一下子就都跑过来了。

然而黑洞表面附近的时空弯曲得太厉害，组成飞船“背影”的光子被束缚、滞留在那里，只能一点点跑出来，越跑越稀，所以远方观测者只能看到飞船的背影越来越暗，飞船越来越慢，越来越发红，最后冻结在黑洞的表面上，消失在那里的黑暗中，却永远看不见飞船落入黑洞。

飞船上的宇航员感觉自己正常地进入黑洞，但他收不到来自前方的信息，因为他已进入黑洞内的单向膜区，只有他的信息可以飞向奇点，奇点那边的信息却根本过不来，也就是说，他看不见奇点。他只是感觉到潮汐力越来越大。

什么叫潮汐力呢？就是造成地球上海洋涨潮落潮的那种力。潮汐力实际上是万有引力的差。

我站在地球表面上，我的重量就是我受到的万有引力。但是我头顶受到的万有引力和脚底受到的万有引力有一个差，这是因为我头顶到地心的距离和脚底到地心的距离有一个差。这个差就是我的身高。这个距离差造成的万有引力差大约是三滴水的重量，它就是我受到的潮汐力，此力太小，我们一般注意不到。

但是月球对地球造成的潮汐力却不能忽略（图4-9）。图中地球的实线表示地球固体部分的表面，虚线表示海面。A点到月球的距离和B点到月球的距离差了一个地球直径，这可

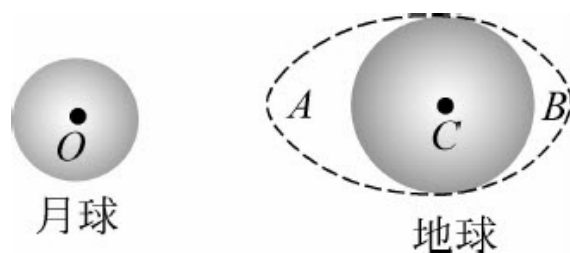


图4-9 月球对地球的潮汐力

是个不能忽略的距离。这使得A处海水受到的月球引力远大于B处海水，这一引力差造成A、B两处海水涨潮，而垂直于月地连线的方向海

水落潮。当然，太阳的引力也对涨落潮有影响，但其影响远小于月球。

进入黑洞的飞船，受到的来自奇点处的巨大潮汐力，远比造成地球上涨落潮的力要大。飞船越接近奇点，受到的潮汐力越大，最终把飞船和宇航员一起撕碎，压入体积为零的奇点。那里是时间终结的地方，飞船和宇航员最后就葬身于那里，并处于“时间之外”。

史瓦西坐标的缺点

式(4.5)所示的史瓦西时空在视界 $r=2M$ （取 $c=G=1$ 的自然单位制）处，存在坐标奇异性，此奇异性把史瓦西时空分成两个部分，洞内和洞外。这两部分各自用一个史瓦西坐标系描写，一个适用于 $r<2M$ ，另一个适用于 $r>2M$ ，但哪一个都不适用于 $r=2M$ 的视界。因此，这两个坐标系是不连通的。

如果用史瓦西坐标来描述一个向黑洞下落的质点（如飞船），或者一束射向黑洞的光。研究表明，质点和光趋近黑洞表面 $r=2M$ 时，将会有

$$t \sim \lim_{r \rightarrow 2M} [-2M \ln(r - 2M)] \rightarrow \infty \quad (4.16)$$

相对论认为，史瓦西时间 t 就是静止于无穷远的观测者的真实时间（图4-10）。所以，在无穷远观测者看来，任何质点和光都不能在有限时间内落到黑洞表面，更不用说进入黑洞内部了。

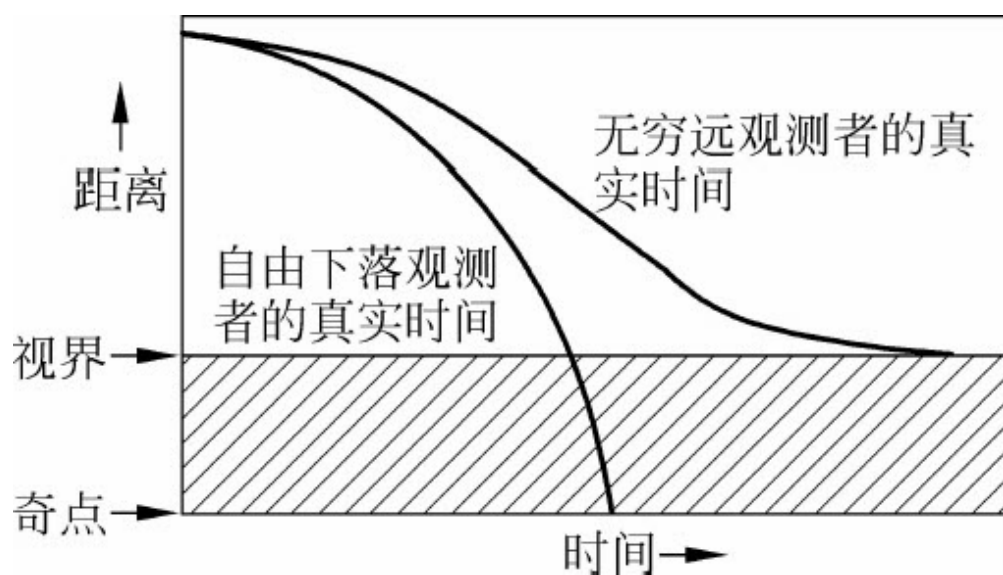


图4-10 黑洞的两种时间

自由下落坐标系

苏联物理学家诺维科夫经过复杂的坐标变换

$$\left\{ \begin{array}{l} \tau = \frac{r_0}{2} \left(\frac{r_0}{2M} \right)^{1/2} (\eta + \sin \eta) \\ R = \left(\frac{r_0}{2M} - 1 \right)^{1/2} \end{array} \right. \quad (4.17)$$

$$\left\{ \begin{array}{l} t = 2M \ln \left| \frac{\left(\frac{r_0}{2M} - 1 \right)^{1/2} + \tan \frac{\eta}{2}}{\left(\frac{r_0}{2M} - 1 \right)^{1/2} - \tan \frac{\eta}{2}} \right| + 2M \left(\frac{r_0}{2M} - 1 \right)^{1/2} \left[\eta + \frac{r_0}{4M} (\eta + \sin \eta) \right] \\ r = \frac{r_0}{2} (1 + \cos \eta) \end{array} \right. \quad (4.18)$$

得到一个自由下落坐标系，称为诺维科夫坐标系，其时空线元为

$$ds^2 = -d\tau^2 + \left(\frac{R^2 + 1}{R^2} \right) \left(\frac{r}{R} \right)^2 dR^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2 \quad (4.19)$$

此坐标系从黑洞外 $r=r_0$ 处，向黑洞自由下落，式中 (τ, R) 为此系的时间与空间坐标。 (η, r_0) 为参数坐标，引入它们是为了使 (τ, R) 与史瓦西坐标 (τ, r) 间的关联表达式不至于太复杂。

研究表明，诺维科夫坐标系的时间 τ 就是自由下落观测者的真实时间（图4-10）。计算表明，这个自由下落观测者将在有限的真实时间 τ 内，穿过黑洞表面（视界）并很快抵达奇点。

显然，这个坐标系可以同时覆盖黑洞外部和内部，以及视界 $r=2M$ 本身，具有一定的优越性。但它还不能最大限度地覆盖整个史瓦西时空。事实上，史瓦西时空的范围要比人们原来想象的更广阔。

史瓦西时空的最大扩展——克鲁斯卡坐标

下面介绍一个能够覆盖整个史瓦西时空的坐标系，克鲁斯卡（Kruskal）坐标系。坐标变换

$$\begin{cases} T = 4M \left(\frac{r}{2M} - 1 \right)^{1/2} e^{r/4M} \operatorname{sh} \frac{t}{4M} \\ R = 4M \left(\frac{r}{2M} - 1 \right)^{1/2} e^{r/4M} \operatorname{ch} \frac{t}{4M} \end{cases} \quad (r > 2M, \text{I 区}) \quad (4.20)$$

$$\begin{cases} T = -4M \left(\frac{r}{2M} - 1 \right)^{1/2} e^{r/4M} \operatorname{sh} \frac{t}{4M} \\ R = -4M \left(\frac{r}{2M} - 1 \right)^{1/2} e^{r/4M} \operatorname{ch} \frac{t}{4M} \end{cases} \quad (r > 2M, \text{II 区}) \quad (4.21)$$

$$\begin{cases} T = 4M \left(1 - \frac{r}{2M} \right)^{1/2} e^{r/4M} \operatorname{ch} \frac{t}{4M} \\ R = 4M \left(1 - \frac{r}{2M} \right)^{1/2} e^{r/4M} \operatorname{sh} \frac{t}{4M} \end{cases} \quad (r < 2M, \text{F 区}) \quad (4.22)$$

$$\begin{cases} T = -4M \left(1 - \frac{r}{2M} \right)^{1/2} e^{r/4M} \operatorname{ch} \frac{t}{4M} \\ R = -4M \left(1 - \frac{r}{2M} \right)^{1/2} e^{r/4M} \operatorname{sh} \frac{t}{4M} \end{cases} \quad (r < 2M, \text{P 区}) \quad (4.23)$$

把史瓦西时空中的线元式（4.5）变换成

$$ds^2 = \frac{2M}{r} e^{-r/2M} (-dT^2 + dR^2) + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \quad (4.24)$$

此式称为克鲁斯卡坐标系下的线元表达式，（R，T）即克鲁斯卡坐标。从式（4.24）的号差可以判定，T是时间坐标，R是空间坐标。其中r与R、T的关系由下式决定：

$$16M^2 \left(\frac{r}{2M} - 1 \right) e^{r/2M} = R^2 - T^2 \quad (4.25)$$

容易看出，克鲁斯卡时空不再是与时间坐标 T 无关的了。然而，它的度规分量在引力半径 r_g 处不再奇异，坐标奇异性被消除了。当然 $r=0$ 的奇点依然存在，内禀奇点不可能通过坐标变换来消除。

克鲁斯卡坐标系可以统一描述整个史瓦西时空，它覆盖了黑洞内、外及视界。而且，从克鲁斯卡时空图4-11可知，它扩大了史瓦西时空。

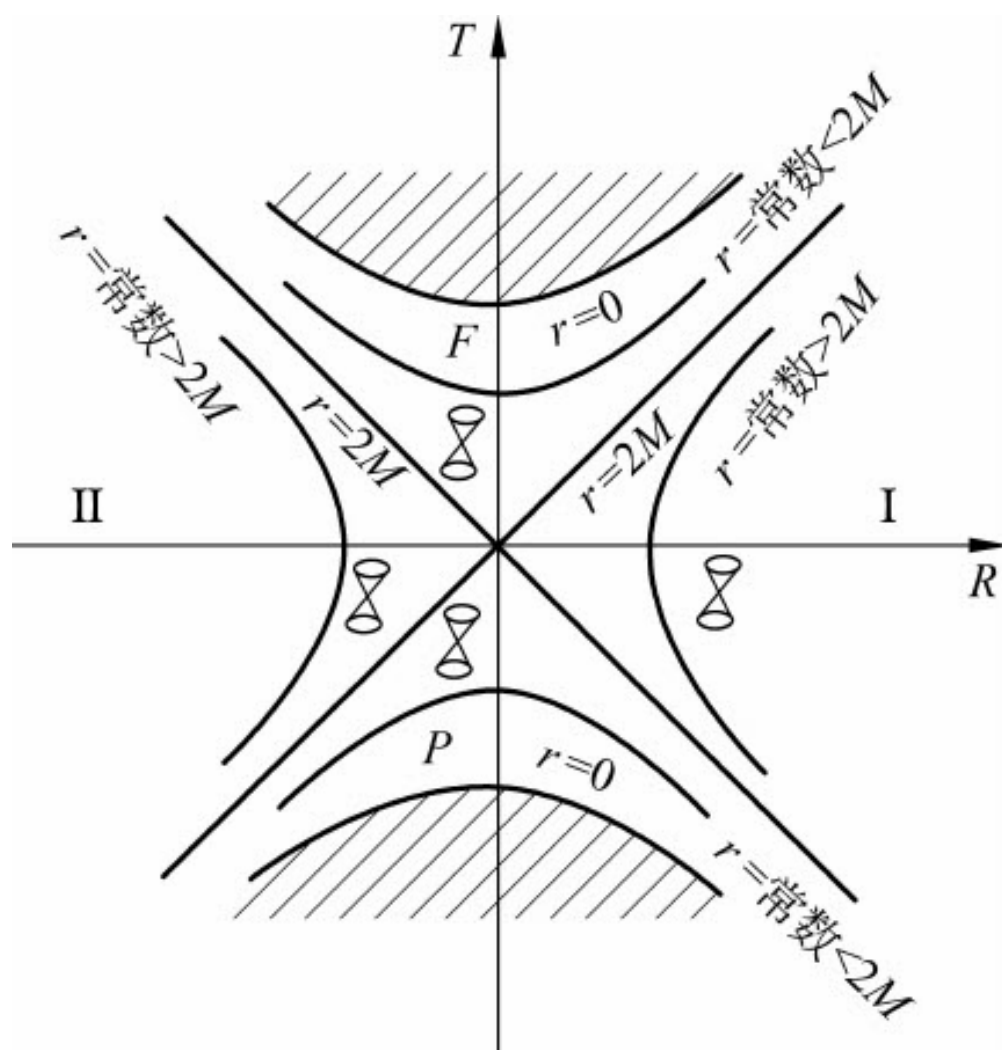


图4-11 克鲁斯卡时空图

图4-11中两条对角线是视界。I区即我们通常谈论的黑洞外部宇宙，F区为黑洞区，P区为白洞区，II区是另一个洞外宇宙，它和我们的

宇宙没有因果连通，没有任何信息交流。这“另一个宇宙”的存在，是我们原来没有想到的。奇点 $r=0$ 分别出现在白洞区和黑洞区，以双曲线形式呈现，阴影区不属于时空。

I 区和 II 区中 $r=\text{常数}$ 的双曲线，就是史瓦西时空中静止粒子的世界线。F区和P区中的 $r=\text{常数}$ 的双曲线为等时线。应当注意，此图中光锥仍呈 45° 角张开，图中的任何一点，都代表一个二维球面。

显然，白洞P中的粒子和信号可进入宇宙 I 和宇宙 II，但宇宙 I、II 中的粒子和信号都不能退回白洞区。I 或 II 中的粒子或信号都可以进入黑洞区F，但 I、II 之间不能交流。黑洞区的粒子和信号也不能倒回宇宙 I 或 II，只能向前到达奇点。

应该指出史瓦西解是场方程的真空解，除去 $r=0$ 外，都是真空区，洞内的单向膜区也不例外，视界处当然也是真空。

在数学上，克鲁斯卡坐标比史瓦西坐标优越，它能覆盖整个史瓦西流形，而且能对流形上的一切过程（黑洞过程、白洞过程等）作最完备的描述，即除去通往内禀奇点的测地线之外，所有的测地线都可以无限延伸，通向无穷远。所以说，克鲁斯卡度规具有最大解析区和最高完备性。

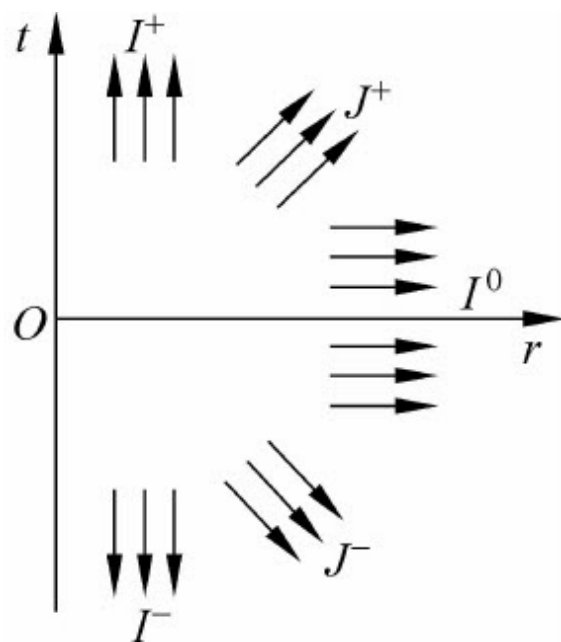
彭罗斯图

为了研究时空的整体性质，特别是在无穷远处的性质，科学家们建议采用共形变换，对时空尺度进行大力压缩，把“无穷远”压到有限的距离之内，使其成为看得见摸得着的东西。这种图就是彭罗斯图。

下面我们先来定义各种无穷远：

- (1) 类时未来无穷远 I^+ ： r 有限， $t \rightarrow +\infty$ ；
- (2) 类时过去无穷远 I^- ： r 有限， $t \rightarrow -\infty$ ；
- (3) 类空无穷远 I^0 ： t 有限， $r \rightarrow \infty$ ；
- (4) 类光未来无穷远 J^+ ： $(t-r)$ 有限， $(t+r) \rightarrow +\infty$ ；
- (5) 类光过去无穷远 J^- ： $(t+r)$ 有限， $(t-r) \rightarrow -\infty$ 。

可以证明，在共形变换下，闵可夫斯基时空图4-12可变成彭罗斯图4-13。同样，克鲁斯卡时空图4-11可变成彭罗斯图4-14。应该注意，共形变换实际上是一种尺度伸缩变换，它把“无穷远”压缩到了有限距离之内。



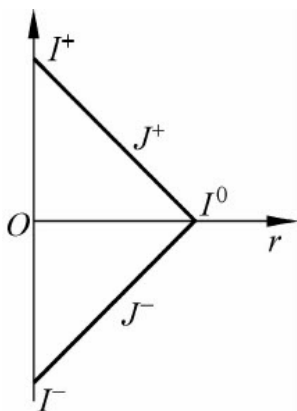


图4-13 闵可夫斯基时空的彭罗斯图

图4-12 闵可夫斯基时空图

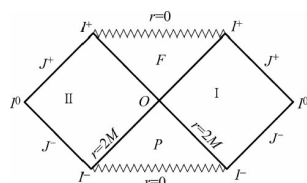


图4-14 克鲁斯卡时空的彭罗斯图

例如，闵可夫斯基时空，它的时间无穷远（类时无穷远）和空间无穷远（类空无穷远），都是既画不出又摸不着的。但利用尺度压缩的共形变换作成彭罗斯图之后，就是既画得出又摸得着的了。 I^0 点就是空间无穷远， I^+ 点和 I^- 点就是时间无穷远，读者不妨用手摸摸这几个无穷远。

应该说明，从我们的上、下、前、后、左、右均可以走向空间无穷远，我们把所有这些空间无穷远“认同”为一个点，就是图中的 I^0 点。同样，我们把所有的时间未来无穷远认同为一个点 I^+ ；所有的时间过去无穷远也认同为一个点 I^- 。 J^+ 和 J^- 两个不含端点（即不含 I^+ 、 I^- 和 I^0 点）的开线段，是类光无穷远。

我们的闵可夫斯基时空，对应不含上述无穷远点（即不含 I^0 、 I^\pm 、 J^\pm ）的开区间。开区间的边界（ I^0 、 I^\pm 、 J^\pm ）实际上并不对应闵可夫斯基时空中的点。也就是说，闵可夫斯基时空不含无穷远点。上述共形变换的尺度压缩，越趋向无穷远，压缩越厉害。在 I^0 、 I^\pm 及 J^\pm 处，尺度产生无限压缩。

还应该注意共形变换的另一特点，它是一个保角变换。虽然时空的尺度发生了压缩，但角度保持不变。原时空中呈 45° 角的光锥，在彭罗斯图中仍保持 45° 角。

图4-14是克鲁斯卡时空的彭罗斯图。此图中任何一点代表一个 r 取定值的二维球面。I、II是两个相互不通信息的宇宙。每个宇宙均有自己的类时无穷远 I^\pm 、类光无穷远 J^\pm 和类空无穷远 I^0 ， $r=2M$ （采用自然单位制）是视界。F是黑洞区，P是白洞区。 $r=0$ 是内禀奇点。奇点和边界（ I^0 、 I^\pm 、 J^\pm ）均不属于克鲁斯卡时空。此时空也是一个开区域，光锥仍保持 45° 角。

爱因斯坦-罗森桥

对于 I 和 II 这两个相互不通信息的宇宙，O 点是一个“喉”，或者叫做“虫洞”，又叫做爱因斯坦-罗森桥。这两个宇宙可以通过此虫洞相通。但是，通过这个虫洞的世界线都是类空的，也就是说，都是超光速物体的世界线。这就是说，只有超光速运动的质点或飞船才能通过此虫洞，从一个宇宙进入另一个宇宙，而超光速运动是相对论理论所禁止的，所以这是一种不可穿越的虫洞。

图4-15是爱因斯坦-罗森桥的镶嵌图。它是这样得到的，在史瓦西时空中，取时间 $t = \text{常数}$ ， $\theta = \frac{\pi}{2}$ ，于是得到一个用 (r, φ) 描述的二维曲面。再引入一维新的空间 z ， z 不是 t 也不是 θ ，是另一维空间。把表示爱因斯坦-罗森桥的曲面 (r, φ) 镶嵌到此三维欧氏空间中，就得到图4-15。

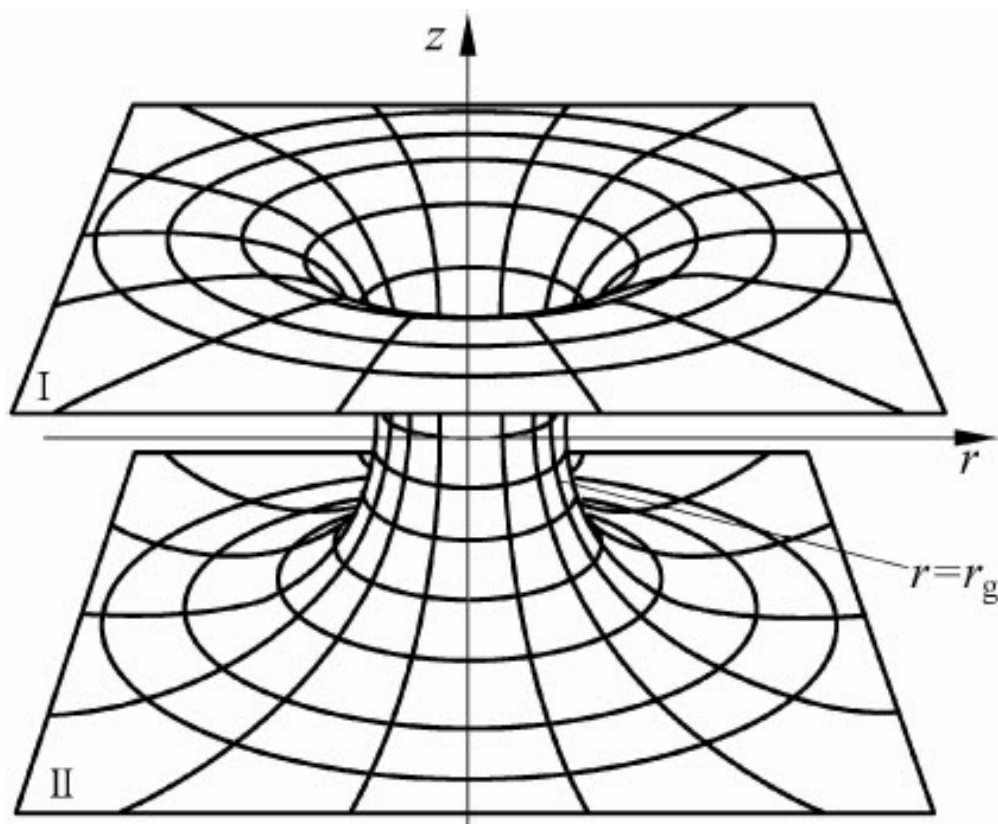


图4-15 爱因斯坦-罗森桥的镶嵌图

注意，图中只有这个曲面属于史瓦西时空，面以外的空间不属于史瓦西时空。曲面上 I、II 两个部分属于图4-11（或图4-14）中互无因果联系的两个宇宙 I 与 II。它们通过 $r=r_g$ 的“喉”相连。

再强调一下，图4-15中，只有曲面属于史瓦西时空的 I、II 两个宇宙，所谓“喉”也只是指 I、II 两部分曲面相连的部分，不包括曲面包围的中空“管道”。I、II 之间可能交流的信号，是沿曲面经 r_g 进入对方的，但都必须是超光速信号，而超光速运动是不可能的，所以 I、II 两部分宇宙互不通信息，更没有物质交流。



绘画: 张京

第五章 黑洞附近的物理效应

带电球体外部的时空

不随时间变化的、带电球状物体周围的时空弯曲状况，可以参照不带电的史瓦西解的求解方式，比较容易地求出。不过这时要用到电磁场方程，把麦克斯韦的电磁方程与爱因斯坦的广义相对论方程联立起来求解，因而也需要克服一些困难。

1916年和1918年Reissner和Nordstrom分别独立地算出了这个解，因此称为Reissner-Nordstrom解（简称R-N解或带电史瓦西解）。在自然单位制下，其时空弯曲情况可以用线元

$$ds^2 = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right) dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2 \quad (5.1)$$

表出，式中M、Q分别为场源的质量和电荷。

这是一个电磁真空解，与史瓦西解不同，这个解由两个参量决定，总质量M和总电荷Q。另外，在带电球体的外面，不是完全的真空，而是存在电磁场。不过，除电磁场外不存在任何其他物质。它所描述的弯曲时空是静态、球对称的，也就是说，时空的弯曲情况呈现球对称，而且不随时间变化。

转动星体外部的时空

1963年，克尔（R.P.Kerr）得出了场方程的一个稳态轴对称解。所谓稳态，就是不随时间变化。此解描述的是一个以恒定角速度转动的星体外部（外部是真空）的时空弯曲情况。那是一个处于真空状态的不随时间变化的、轴对称的弯曲时空。在自然单位制下，它的线元是

$$ds^2 = -\left(1 - \frac{2Mr}{\rho^2}\right)dt^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2 + \left[(r^2 + a^2)\sin^2\theta + \frac{2Mra^2 \sin^4\theta}{\rho^2}\right]d\varphi^2 - \frac{4Mra \sin^2\theta}{\rho^2}dt d\varphi \quad (5.2)$$

其中

$$\rho^2 = r^2 + a^2 \cos^2\theta, \quad \Delta = r^2 - 2Mr + a^2$$

不难看出度规（即坐标微分元前面的系数）不含 t 和 φ 。不含 t 表明时空弯曲情况不随时间变化，因而处于稳态；不含 φ 表明时空弯曲情况在角 φ 的各个方向完全一样，因而是轴对称的。

与史瓦西时空不同，此解也由两个参量决定，即质量 M 和角动量 J 。在式（5.2）中，角动量 J 是以单位质量角动量 a 的形式出现的， $a = J/M$ 。

此解的求出十分艰难，克尔在求解过程中有不少创新。最初许多人看不懂他的求解方法。他最早在一个天体物理研讨会上报告了自己的工作，但没有引起与会者的注意。

那时天文界的兴趣集中在“类星体”上。这是一种奇怪的天体，它红移量很大，按照哈勃定律，它应该离我们非常远，但它又特别亮，这表

明它发光能力极强，用当时已知的物理知识很难解释。与会者一直在热烈讨论类星体，克尔介绍了自己的工作后，没有人理会他，大家仍在议论类星体。一位了解克尔工作的意义的学者站起来发言，提醒大家注意克尔的工作，但仍无人理会，大家继续讨论类星体。

后来学术界终于认识到了克尔解的重大意义，这个解的出现极大地推动了相对论天体物理学的研究，并导致了黑洞研究的高潮。

转动带电星体外部的时空

后来，纽曼（E. T. Newman）等人把克尔解推广到带电情况，得到克尔-纽曼（Kerr-Newman）解，它描述一个转动带电星体的外部引力场，即描述该星体外部时空的弯曲情况。在自然单位制下，其线元为

$$ds^2 = - \left(1 - \frac{2Mr - Q^2}{\rho^2} \right) dt^2 + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 + \left[(r^2 + a^2) \sin^2 \theta + \frac{(2Mr - Q^2) a^2 \sin^4 \theta}{\rho^2} \right] d\varphi^2 - \frac{2(2Mr - Q^2) a \sin^2 \theta}{\rho^2} dt d\varphi \quad (5.3)$$

式中 $\rho^2 = r^2 + a^2 \cos^2 \theta$ ， $\Delta = r^2 - 2Mr + a^2 + Q^2$ 。与克尔解类似，这也是一个稳态轴对称的时空，度规中不含 t 和 φ 。它由三个参数决定，星体的总质量 M 、总角动量 J 和总电荷 Q ，角动量 J 也是以单位质量角动量 a 来显示的， $a = J/M$ 。

注意，上述解都是真空解，即星体外部都是真空区。带电情况下，星体外部是电磁真空，即不存在电磁场以外的任何物质。

克尔-纽曼时空在 $M \neq 0$ ， $J \neq 0$ 但 $Q = 0$ 时回到克尔时空；在 $M \neq 0$ ， $Q \neq 0$ 但 $J = 0$ 时回到 R-N 时空；在 $M \neq 0$ ， $J = 0$ ， $Q = 0$ 时回到史瓦西时空。读者不妨参照式（5.1）～式（5.3），以及式（4.5）自己算一下。研究表明，克尔-纽曼时空是最一般的稳态轴对称时空，因而十分重要。

再谈度规

我们看到式（5.2）和式（5.3）所示的克尔时空与克尔-纽曼时空的度规，都比以前看到的时空度规式（4.3）～式（4.5）以及式（5.1）更为复杂，出现了 $dt d\varphi$ 项。

实际上，度规共有16项（即16个分量），只不过在许多情况下，不少度规的分量为零罢了。

在相对论中，为了表达简洁，弯曲时空中的线元和度规通常简写为

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (\mu, \nu = 0, 1, 2, 3) \quad (5.4)$$

其中，上、下角标各有一个 μ ，还各有一个 ν 。爱因斯坦规定，重复的上下角标表示求和。这是爱因斯坦本人在数学上的一个小创造。以往表示求和时，数学家都会在表达式前面加一个求和号 \sum ，例如

$$ds^2 = \sum_{\mu, \nu} g_{\mu\nu} dx^\mu dx^\nu$$

爱因斯坦发现，只要规定重复角标代表求和， \sum 号就可以省去，不会造成混乱，也不会影响计算。

若是展开来写，式（5.4）的右边会十分复杂，有16项

$$ds^2 = g_{00} (dx^0)^2 + g_{01} dx^0 dx^1 + g_{10} dx^1 dx^0 + g_{11} (dx^1)^2 + \dots$$

式中 $x^0 = ct$ ，表示时间， c 为光速。 x^1 、 x^2 和 x^3 表示3个空间坐标，如 (x, y, z) 或 (r, θ, φ) 。

度规 $g_{\mu\nu}$ 是时空坐标 (x^0, x^1, x^2, x^3) 的函数，它是对称的

$$g_{\mu\nu} = g_{\nu\mu}$$

例如 $g_{01} = g_{10}$ ， $g_{12} = g_{21}$ ，等等。所以，16个度规分量中只有10个独立。

闵可夫斯基时空度规为 $g_{00} = -1$ ， $g_{11} = g_{22} = g_{33} = 1$ ，其余分量（如 g_{01} ， g_{12} 等）为零。

史瓦西度规为

$$g_{00} = -\left(1 - \frac{2GM}{c^2 r}\right), \quad g_{11} = \left(1 - \frac{2GM}{c^2 r}\right)^{-1}, \quad g_{22} = r^2, g_{33} = r^2 \sin^2 \theta \quad (5.5)$$

其余分量为零。

在自然单位制（ $c = G = \hbar = 1$ ）下，克尔-纽曼时空度规为

$$\left\{ \begin{array}{l} g_{00} = -\left(1 - \frac{2Mr - Q^2}{\rho^2}\right) = -\frac{(r - r_+)(r - r_-)}{\rho^2} \\ g_{11} = \frac{\rho^2}{\Delta} = \frac{\rho^2}{(r - r_+)(r - r_-)} \\ g_{22} = \rho^2 \\ g_{33} = \left[(r^2 + a^2) \sin^2 \theta + \frac{(2Mr - Q^2)a^2 \sin^4 \theta}{\rho^2}\right] \\ g_{03} = g_{30} = -\frac{(2Mr - Q^2)a \sin^2 \theta}{\rho^2} \end{array} \right. \quad (5.6)$$

其中 $\rho^2 = r^2 + a^2 \cos^2 \theta$ ， $\Delta = r^2 - 2Mr + a^2$ 。 r_+ 、 r_- 、 r_+^s 、 r_-^s 的意义可参见下面式（5.10）及式（5.17），度规的其余分量为零。

前面已经讲过，知道了度规，就知道了时空的弯曲情况。求解广义相对论中的爱因斯坦方程（即场方程），就是要求出全部度规函数（共10个独立分量）。场方程中表示时空曲率的量 $R_{\mu\nu}$ 和 R 是度规及其一阶导数和二阶导数的函数，所以场方程是关于度规的二阶微分方程，而且是非线性方程，没有通行的解法，求解过程十分困难、复杂。每求出一个解，都可以用求解者的名字命名。

时轴正交，静态与稳态

度规中两个下标不同的分量（例如 g_{01} 、 g_{12} 、...），表示相关的两个坐标轴是否正交（垂直）。例如，上式中 $g_{01}=0$ ，表示时间轴（ x^0 ，即 t 轴）与空间的 x^1 轴（此处即 r 轴）正交； $g_{12}=0$ ，表示 x^1 与 x^2 两个空间轴（即 r ， θ 轴）正交。 $g_{03}\neq 0$ 则表示时间轴与 x^3 轴（即 ϕ 轴）不正交，在物理上这反映了作为引力源的物体在绕 ϕ 轴转动，整个时空也被这一转动所拖动。

不难看出，平直的闵可夫斯基时空和弯曲的史瓦西时空，所有的四个轴都是正交的，克尔时空和克尔-纽曼时空，由于 $g_{03}\neq 0$ ，时间轴和一个空间轴不正交。

时间轴和3个空间轴都正交的时空（即 $g_{01}=g_{02}=g_{03}=0$ 的时空），称为“时轴正交”的时空。

前面已经谈到，如果度规函数不含时间 t （即 x^0 ），时空就不随时间变化，这样的时空称为稳态时空。如果稳态时空时轴正交，就称为静态时空。显然，静态时空一定稳态，而稳态时空不一定静态。闵可夫斯基时空、史瓦西时空和R-N时空都是静态时空，而克尔时空和克尔-纽曼时空则是稳态时空。

奇异性

从式（5.3）不难看出，克尔-纽曼时空在

$$\rho^2 = r^2 + a^2 \cos^2 \theta = 0 \quad (5.7)$$

和

$$\Delta = r^2 - 2Mr + a^2 + Q^2 = 0 \quad (5.8)$$

处，存在奇异性。即在这两处度规的某些值会出现无穷大。研究表明，式（5.7）所示的奇异性的位置出现在

$$\begin{cases} r = 0 \\ \theta = \pi/2 \end{cases} \quad (5.9)$$

处，呈现为一个奇环。这是一种内禀奇异性，时空曲率为无穷大，而且不能通过坐标变换来消除。式（5.8）所示的奇异性位于

$$r_{\pm} = M \pm \sqrt{M^2 - a^2 - Q^2} \quad (5.10)$$

处，呈现为两个椭球面。这是一种坐标奇异性。注意，因为采用了自然单位制，所以本节的公式才显得比较简洁。如果恢复到普通单位时，式（5.10）则写成

$$r_{\pm} = \frac{GM}{c^2} \pm \sqrt{\left(\frac{GM}{c^2}\right)^2 - \left(\frac{J}{Mc}\right)^2 - \frac{GQ^2}{c^4}} \quad (5.11)$$

需要注意的是，恢复到普通单位时，单位质量角动量 \mathbf{a} 应改写为 $\mathbf{a} = \mathbf{J}/(Mc)$ 。

上述三种时空中都存在黑洞，式（5.10）给出的坐标奇异性出现的地方，就是黑洞的事件视界。

带电的黑洞

当带电史瓦西时空中的物质和电荷聚集到星体中心时，就会出现黑洞，称为Reissner-Nordstrom黑洞，或带电史瓦西黑洞（图5-1）。

黑洞的中心 $r=0$ 处，存在奇点。这种黑洞有两个视界和两个无限红移面。外视界 r_+ 与外无限红移面重合，内视界 r_- 与内无限红移面重合，分别位于

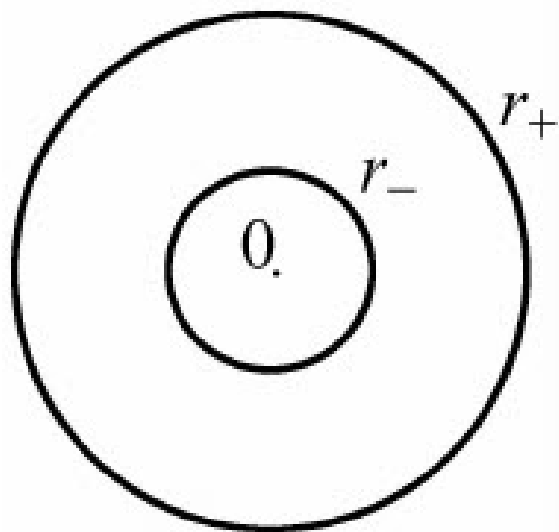


图5-1 带电史瓦西黑洞的结构

$$r_{\pm} = \frac{GM}{c^2} \pm \sqrt{\left(\frac{GM}{c^2}\right)^2 - \frac{GQ^2}{c^4}} \quad (5.12)$$

处，式中 M 、 Q 分别为黑洞的质量和电荷， G 、 c 分别为万有引力常数和光速。当使用自然单位制（ $G=c=\hbar=1$ ）时，上式简化为

$$r_{\pm} = M \pm \sqrt{M^2 - Q^2} \quad (5.13)$$

显然，它们都是球面。

这种黑洞的单向膜区位于内、外视界之间。在那个区域 r 表示时间， t 表示空间。从外部穿过外视界进入单向膜区的物质将不能停留，必须落向内视界，并穿过它进入 $r < r_-$ 的区域。应该注意， $r < r_-$ 的区域不是单向膜区， t 恢复为时间， r 恢复为空间，进入那里的物质不一定趋向奇点。

相反，研究表明， $r=0$ 处的奇点（现在是球心）似乎表现出斥力，拒绝让任何物质靠近它。只有光可以到达这个奇点。如果进入洞中的火箭想靠近奇点，则必须加速，而且只有加速度趋于无穷大时才能到达奇点。

旋转的黑洞

当克尔时空中的物质都聚向中心时，也会出现黑洞，称为克尔黑洞。克尔黑洞也有两个视界

$$r_{\pm} = \frac{GM}{c^2} \pm \sqrt{\left(\frac{GM}{c^2}\right)^2 - \left(\frac{J}{Mc}\right)^2} \quad (5.14)$$

式中M、J分别为黑洞的质量和角动量。当使用自然单位制时，上式化简为

$$r_{\pm} = M \pm \sqrt{M^2 - a^2} \quad (5.15)$$

其中

$$a = J / (Mc)$$

为单位质量的角动量。这种黑洞也有两个无限红移面，而且无限红移面和视界分开了，内、外无限红移面分别是

$$r_{\pm}^* = M \pm \sqrt{M^2 - a^2 \cos^2 \theta} \quad (5.16)$$

这种黑洞的结构如图5-2所示。最外面是外无限红移面，好像一个橘子。外视界犹如一个大核桃，里面套个小核桃，那就是内视界。再往里是内无限红移面。

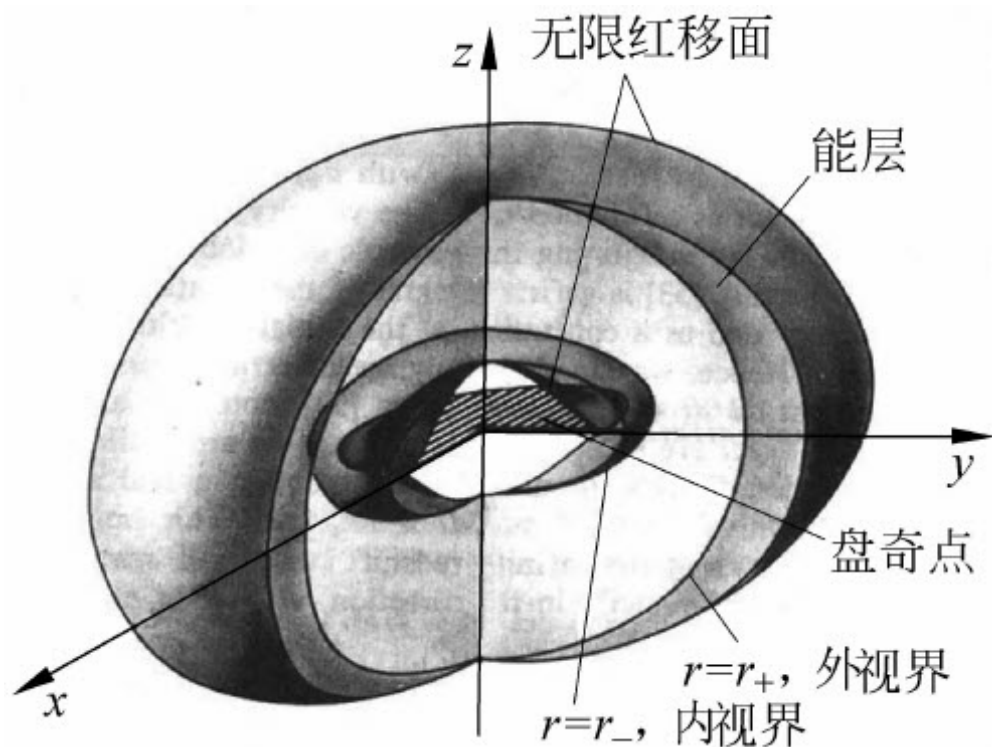


图5-2 克尔黑洞的剖面图

从式 (5.14) 或式 (5.15) 看, r_+ 和 r_- 似乎均与 θ 、 ϕ 角无关, 内、外视界好像都是球面。实际上不是这样, 内、外视界其实都是轴对称的椭球。

造成这一误解的原因是, 我们用来描述克尔黑洞的坐标不是球坐标, 而是椭球坐标。

视界与无限红移面之间是能层, 在 r_+ 与 r_+^s 之间是外能层, 在 r_- 与 r_-^s 之间是内能层。能层只是黑洞的附属物。穿过外无限红移面进入外能层的飞船还不算进入了黑洞, 它仍可以逃出去。克尔黑洞的边界是外视界, 穿过外视界才算进入了黑洞, 就逃不出去了。

单向膜区位于外视界 r_+ 与内视界 r_- 之间, 那里是时空坐标互换区

（即单向膜区），进入那里的飞船或物质将不能停留，必须穿过内视界进入 $r < r_-$ 的区域。那里不再是时空坐标互换区，飞船可以在那里自由飞行。

注意，克尔黑洞的中心不是奇点，而是奇环。在图5-2中，读者可以看到一个用阴影表示的盘，盘内部并不奇异，只有它的边缘是奇异的，那就是奇环。奇环不能简单地用 $r=0$ 来表示， $r=0$ 表示的是整个用阴影表出的盘，再加上 $\theta = \frac{\pi}{2}$ 的限制才表示盘的边缘，那就是奇环。读者可能感到很有趣， $r=0$ 不表示球心，而表示一个圆盘，这也是由于我们采用的坐标不是普通的球坐标，而是椭球坐标的缘故。

奇环位于

$$\begin{cases} r = 0 \\ \theta = \pi/2 \end{cases}$$

处，这正是式（5.9）给出的内禀奇异性的位置。

旋转、带电的黑洞

很容易把上述工作拓广到旋转、带电的黑洞，这种黑洞称为克尔-纽曼黑洞，或带电克尔黑洞。克尔-纽曼黑洞的结构与克尔黑洞非常相似，有两个视界

$$r_{\pm} = M \pm \sqrt{M^2 - a^2 - Q^2}$$

此即式（5.10）。用普通单位制表出则为

$$r_{\pm} = \frac{GM}{c^2} \pm \sqrt{\left(\frac{GM}{c^2}\right)^2 - \left(\frac{J}{Mc}\right)^2 - \frac{GQ^2}{c^4}}$$

此即式（5.11）。这种黑洞也有两个无限红移面

$$r_{\pm}^s = M \pm \sqrt{M^2 - a^2 \cos^2 \theta - Q^2} \quad (5.17)$$

它们都互不重合。克尔-纽曼黑洞也有内、外两个能层，中心也是一个奇环。也位于式（5.9）

$$\begin{cases} r = 0 \\ \theta = \pi/2 \end{cases}$$

所示的内禀奇异性位置。

二者的差别是，克尔黑洞的奇环与内无限红移面接触（图5-3左），克尔-纽曼黑洞的奇环则不与内无限红移面接触（图5-3右）。这是因为克尔-纽曼奇环带电的缘故。

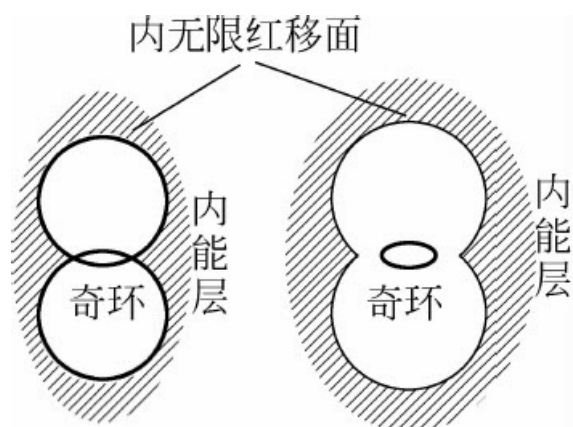


图5-3 奇环与内无限红移面

单向膜区与能层

图5-2是克尔黑洞的剖面图（克尔-纽曼黑洞的剖面图与此类似），两个视界之间的区域是单向膜区。外视界 r_+ 与外无限红移面 r_+^s 之间的空间称为外能层，它实际上位于黑洞之外。内视界 r_- 和内无限红移面 r_-^s 之间的区域称内能层。图中能层区用阴影部分显示出来，下面简述各区的时空特征。为此，我们先把式（5.3）所示的克尔-纽曼时空度规简写为

$$ds^2 = g_{00}dt^2 + 2g_{03}dtd\varphi + g_{11}dr^2 + g_{22}d\theta^2 + g_{33}d\varphi^2 \quad (5.18)$$

其中各度规分量如式（5.6）所示。

现在我们来试图判断一下克尔-纽曼黑洞各时空区的时间坐标和空间坐标。

（1）外无限红移面之外（ $r > r_+^s$ ）：

$g_{00} < 0$ ， $g_{11} > 0$ ， t 表示时间， r 表示空间；

（2）外能层区（ $r_+^s > r > r_+$ ）：

$g_{00} > 0$ ， $g_{11} > 0$ ，时空概念看不清；

（3）单向膜区（ $r_+ > r > r_-$ ）：

$g_{00} > 0$ ， $g_{11} < 0$ ， t 表示空间， r 表示时间，时空坐标互换；

(4) 内能层区 ($r_- > r > r_-^s$) :

$g_{00} > 0$, $g_{11} > 0$, 时空概念看不清;

(5) 内能层以内 ($r_-^s > r$) :

$g_{00} < 0$, $g_{11} > 0$, t表示时间, r表示空间。

我们看到, 内无限红移面包围着一块与洞外宇宙时空属性相同的区域, t表示时间, r表示空间。在单向膜区, 时空坐标互换, t表示空间, r表示时间。但在两个能层区, g_{00} 与 g_{11} 同号, 时空概念看不清楚。

如果我们采用拖动系, 即考虑到任何物理的坐标系都不可避免地转动着的球体所拖动, 假设拖动角速度为

$$\dot{\varphi} = \frac{d\varphi}{dt} = -\frac{g_{03}}{g_{33}} \quad (5.19)$$

则线元式 (5.18) 将化为

$$\begin{aligned} ds^2 &= \left(g_{00} - \frac{g_{03}^2}{g_{33}} \right) dt^2 + g_{11} dr^2 + g_{22} d\theta^2 + g_{33} \left(d\varphi + \frac{g_{03}}{g_{33}} dt \right)^2 \\ &= \hat{g}_{00} dt^2 + g_{11} dr^2 + g_{22} d\theta^2 \end{aligned} \quad (5.20)$$

式中

$$\begin{aligned} \hat{g}_{00} &= g_{00} - \frac{g_{03}^2}{g_{33}} = \frac{-\rho^2 \Delta}{(r^2 + a^2)^2 - \Delta a^2 \sin^2 \theta} = \frac{-\rho^2 (r - r_+)(r - r_-)}{(r^2 + a^2)\rho^2 + (2Mr - Q^2)a^2 \sin^2 \theta} \\ g_{11} &= \frac{\rho^2}{(r - r_+)(r - r_-)} \end{aligned} \quad (5.21)$$

在能层中 $\hat{g}_{00} < 0$, $g_{11} > 0$, 此时t仍为时间, r仍为空间, 能层中时空的

概念可以看清楚了。

下面我们将证明，能层内的观测者和粒子一定会被引力场拖动。这是因为能层内静止的观测者和粒子将是超光速的，所以能层内不可能存在静止的物体，这也是无限红移面又被称为静界的原因。静界就是存在静止观测者和粒子的界限，静界以外可以存在静止观测者和静止粒子，但静界以内不行。

综上所述，对于克尔-纽曼黑洞，时空坐标互换的单向膜区，仅存在于内、外视界之间。在外视界以外，内视界以内， t 都表示时间， r 都表示空间。

拖曳效应与静界

下面我们就来介绍，为什么能层中的质点不能处于静止状态，一定会被转动的黑洞拖动，与黑洞一起旋转，这种效应叫拖曳效应。

现在考虑克尔-纽曼时空中的一个静止质点，静止意味着 $dr=d\theta=d\phi=0$ ，从式（5.3）可以看出，这时线元简化为

$$ds^2 = -\left(1 - \frac{2Mr - Q^2}{\rho^2}\right) dt^2 = -\frac{(r - r_+^s)(r - r_-^s)}{\rho^2} dt^2 \quad (5.22)$$

其中 r_+^s 与 r_-^s 分别为式（5.17）所示的内外无限红移面。在 $r > r_+^s$ 或 $r < r_-^s$ 时， $ds^2 < 0$ ，这表示静止在外无限红移面以外，或内无限红移面以里的质点，时空间隔为负，描出的世界线类时，处于亚光速运动状态，这是相对论允许的。所以，质点可以在内无限红移面以里，或外无限红移面之外的广大时空区静止。

在 $r_+ < r < r_+^s$ 的外能层区，以及 $r_-^s < r < r_-$ 的内能层区，从式（5.22）不难看出，这时静止质点的 $ds^2 > 0$ ，它表示时空间隔为正（类空），静止质点的世界线类空，该静止质点实际上处在超光速运动状态，这是相对论所不允许的。

当 $r_- < r < r_+$ 时，即质点处于内外视界之间的单向膜区中，此区中时空坐标互换， t 表示空间， r 表示时间，静止质点不能用式（5.22）描述。我们前面已经谈到，这时质点不能静止，必定穿过单向膜区和内视界，进入内能层区。

那么，质点可不可以在无限红移面上静止呢？从式（5.22）不难看

出，当 $r=r_+^s$ 或 $r=r_-^s$ 时，都有 $ds^2=0$ ，间隔为零，世界线类光，它表示这样的质点将以光速运动，这也是相对论所不允许的。

所以质点不能在能层（包括外能层和内能层）中静止，也不能在无限红移面上静止，只能在外无限红移面之外（ $r>r_+^s$ ）或内无限红移面以内（ $r<r_-^s$ ）的广大区域内静止。

所以，我们说无限红移面是质点静止的界线，因此黑洞的无限红移面又称“静界”。

能层中的质点，不能静止，一定会被转动的黑洞“拖动”。有人认为，这种拖曳效应，正是马赫当年预言的效应。马赫早在相对论诞生之前，就猜测转动物体（或作加速运动的物体）会对外部物体产生拖动，并认为这一拖动是惯性效应的根源。当然，这种说法并未得到一致的公认。

可以证明，能层中质点被拖曳的速度

$$\Omega = \frac{d\varphi}{dt}$$

有一个范围，如下式所示：

$$\frac{-g_{03} - \sqrt{-\hat{g}_{00}g_{33}}}{g_{33}} \leq \Omega \leq \frac{-g_{03} + \sqrt{-\hat{g}_{00}g_{33}}}{g_{33}} \quad (5.23)$$

质点转动的角速度就被限制在上述范围之内。由于在内、外视界上 $\hat{g}_{00}=0$ ，从式（5.23）不难看出内、外视界上的拖曳速度必须取确定值，可以算出为

$$\Omega_{\pm} = \lim_{r \rightarrow r_{\pm}} \left(-\frac{g_{03}}{g_{33}} \right) = \frac{a}{r_{\pm} + a^2}$$

人们把外视界的转动角速度 Ω_+ 称为黑洞的转动角速度，通常用 Ω_H 表示

$$\Omega_H = \frac{a}{r_H^2 + a^2} \quad (5.24)$$

其中 H 表示外视界。

不过，应该注意，此速度恰为光速，所以质点不可能在视界面上静止，可把 Ω_H 看作沿视界面运动的光线的速度。

克尔-纽曼时空的彭罗斯图

图5-4给出了克尔-纽曼时空的彭罗斯图，图中没有画出无限红移面， r_+ 为外视界， r_- 为内视界。I区是洞外宇宙，II区是单向膜区，III区是内视界以里的区域。

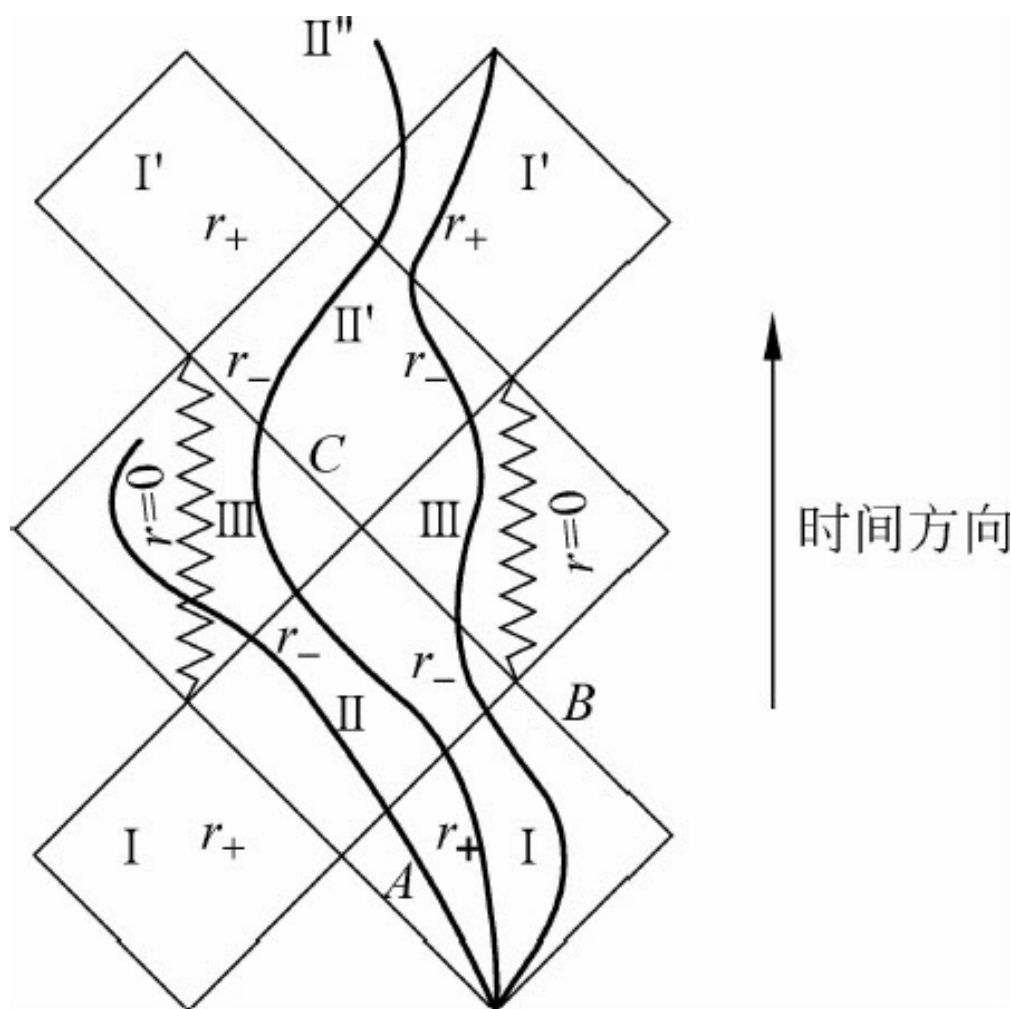


图5-4 克尔-纽曼时空的彭罗斯图

在前面讨论的史瓦西情况下，奇点 $r=0$ 附近， $g_{00}>0$ ， $g_{11}<0$ ， t 是空间坐标， r 是时间坐标。 $r=$ 常数的面是等时面， $r=0$ 的“面”（即奇点）在彭罗斯图上应该与时间方向垂直，这种奇异性称为“类空奇异

性”，它将阻断所有进入黑洞的质点或光子的“前程”。这些粒子都将不可避免地撞在此类空奇点上。所有进入史瓦西黑洞的类时、类光世界线都将在 $r=0$ 的奇点处终止（参看图4-14）。

但是，克尔-纽曼黑洞与史瓦西黑洞不同，在 $r=0$ 附近， $\hat{g}_{00}<0$ ， $g_{11}>0$ ， t 仍是时间坐标， r 是空间坐标，等 r 面与时间方向平行， $r=0$ 作为一个特殊的等 r “面”当然也不例外。与时间走向平行的奇点（或奇环）称为“类时奇点（或奇环）”，所以克尔-纽曼黑洞的奇异性（包括克尔黑洞、克尔-纽曼黑洞的奇环和R-N黑洞的奇点）是类时的。类时奇点（或奇环）现在是竖直的，它不阻挡进入黑洞的光和质点的世界线，除非它们有意往奇点上撞。

进入克尔-纽曼黑洞的粒子，在穿过单向膜区II，进入III区后，还可从另一个区域II'（白洞的单向膜区）中穿过，进入另一个宇宙I'。在I区看来，这个洞是黑洞，II是黑洞的单向膜区，III是黑洞的内部区。而在I'区看来，II'区是白洞的单向膜区，III是白洞内部区。上述彭罗斯图往上、下方可无限延伸，从一个宇宙进入克尔-纽曼黑洞的飞船，有可能从另一个宇宙的白洞中冒出来。

这就是说，两个宇宙之间存在一个虫洞（隧道），黑洞是虫洞的入口，白洞是出口。也有可能I和I'是同一个宇宙，即虫洞的两个开口（白洞和黑洞）均在同一个宇宙中。

世界线A所示的观测者进入黑洞后，穿过奇环进入 $-\infty<r<0$ 的反引力宇宙，此宇宙没有视界，也没有无限红移面。但对于R-N黑洞， $r=0$ 处是奇点，不是奇环，不存在反引力宇宙，此路不通，世界线在 $r=0$ 处终结。世界线B所示的观测者进入黑洞后，又从白洞出来，到达另一个引力宇宙，直到无穷远。世界线C所示的观测者，从黑洞进去白洞出

来，到达另一个宇宙后，又进入那个宇宙的一个黑洞。

不过，真实的黑洞可能不是这样。下面的宇宙监督假设告诉我们，黑洞内部的结构不稳定，稍加扰动，类时奇异性将倒在内视界 r_+ 上，通道就会被阻断，使进入黑洞的飞船不可能从内部穿过，再从一个白洞钻出来，因此黑洞内部并不存在可穿越的时空隧道。

极端黑洞和裸奇异

从式（5.13）可以看出，当增加R-N黑洞的电荷时， r_+ 与 r_- 间的距离会缩小，也就是说单向膜区会变薄。当电荷增加到一定程度，使得 $Q^2 = M^2$ 时，将会有

$$r_+ = r_- = M$$

也就是说，单向膜区缩成一个薄膜，这种黑洞称为极端黑洞，如果再增加一点电荷，则会有 $Q^2 > M^2$ ，这时 r_+ 与 r_- 将会成为复数，这表示单向膜区消失了，奇点会裸露在外面，如图5-5所示。

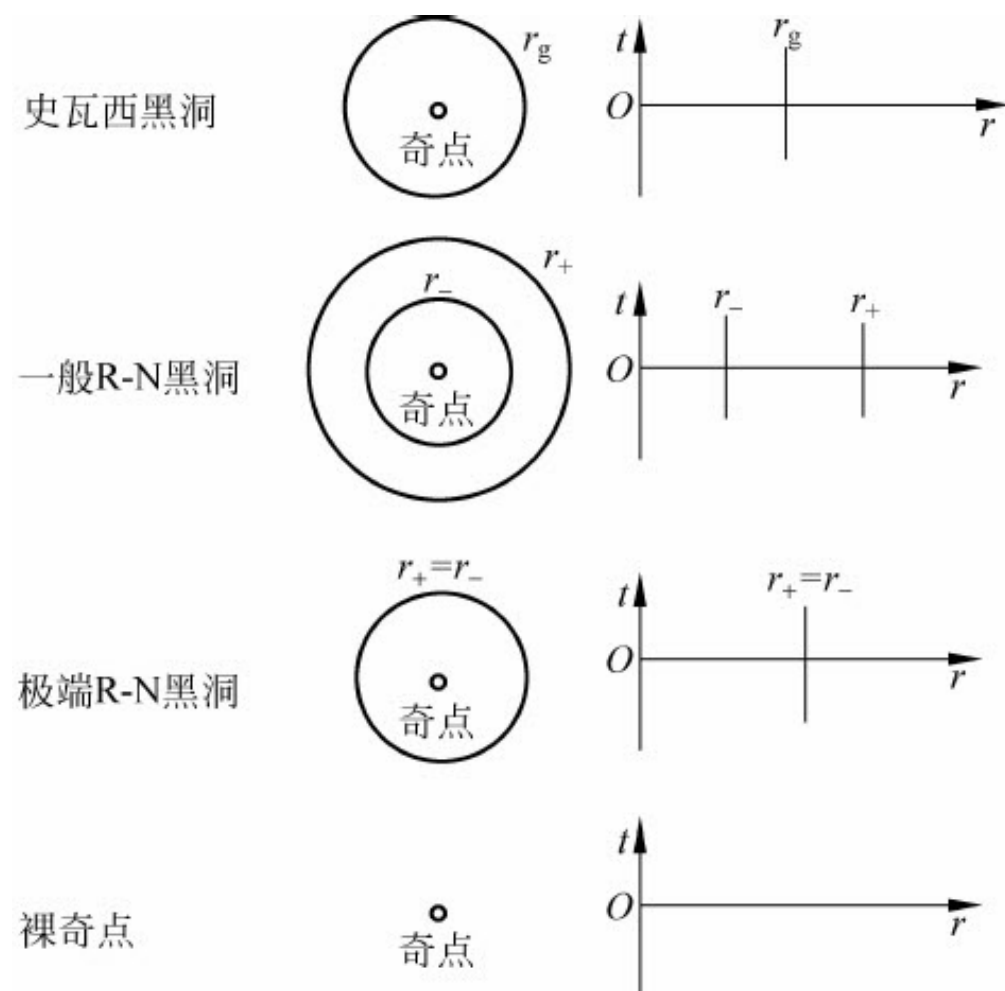


图5-5 R-N黑洞的各种情况

从式 (5.15) 则可看出，当增加克尔黑洞的角动量 J 时， r_+ 也会与 r_- 靠近，单向膜区也会变薄。当 J 增加到使 $a^2 = M^2$ 时，单向膜区也会缩成一个薄膜， $r_+ = r_- = M$ ，克尔黑洞也会成为极端黑洞（图5-6）。这时如果再增加一点角动量，将会有 $a^2 > M^2$ ，单向膜区也会消失，于是奇环会裸露出来，成为裸奇环。我们把裸奇点和裸奇环统称为裸奇异（或裸奇点），如图5-7所示。注意，图5-7未画出无限红移面。

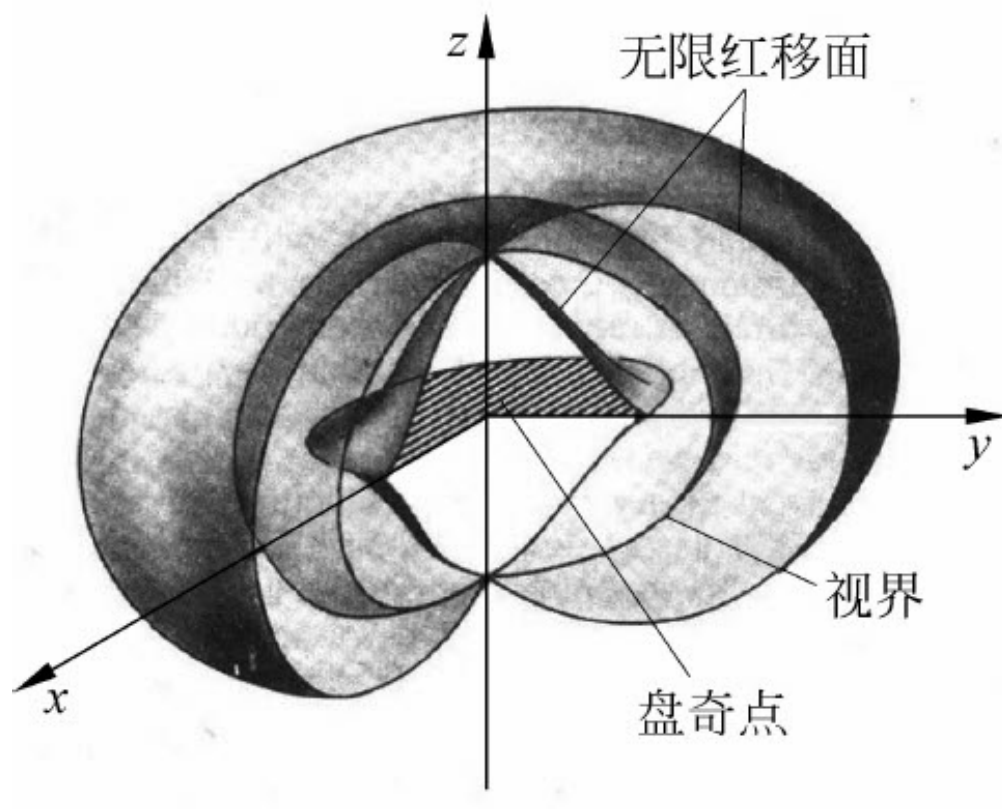


图5-6 极端克尔黑洞，内外视界合在一起

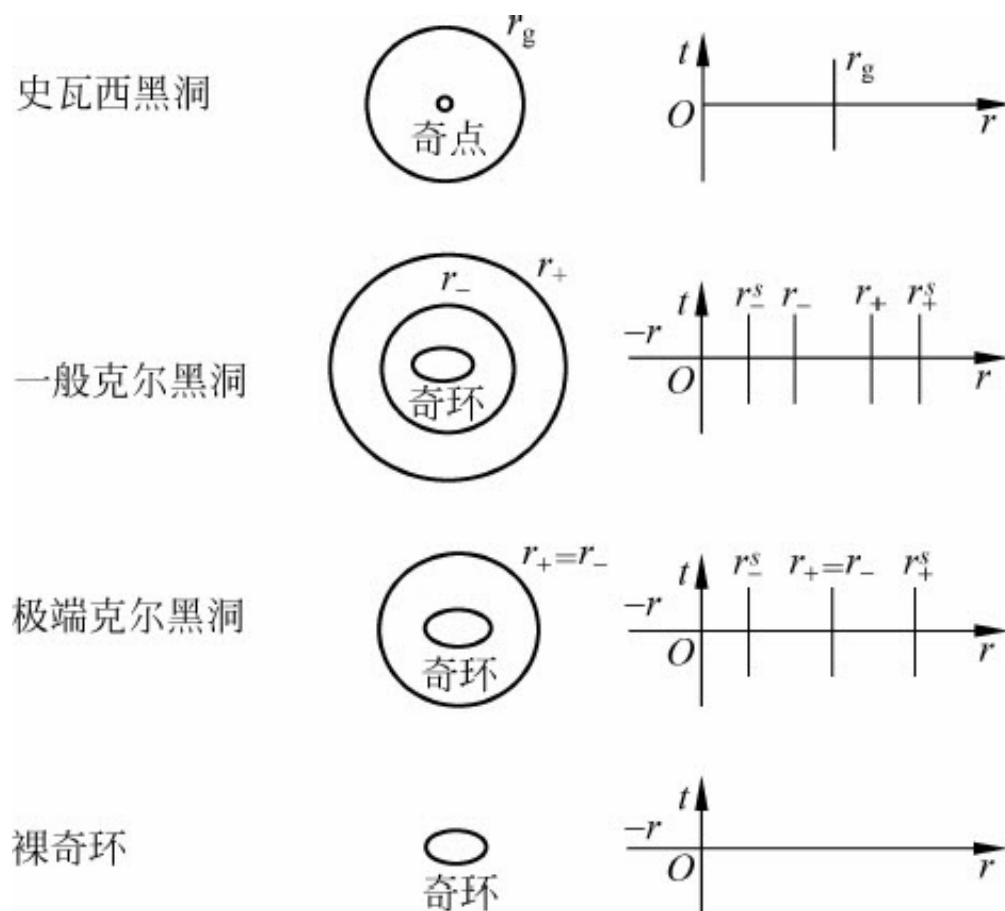


图5-7 克尔黑洞的各种情况

从式 (5.10) 也可看出, 对于克尔-纽曼黑洞, 如果增加电荷 Q , 或者增加角动量 J , 或者同时增加 J 和 Q , 都会使单向膜区变薄, 当 $a^2 + Q^2 = M^2$ 时, 也会形成极端黑洞, 这时单向膜区成为一层薄膜, 再增加 J 或 Q , 同样可以使单向膜区消失, 使奇环裸露出来, 形成裸奇异。

宇宙监督假设

研究表明，奇点或奇环裸露出来是很不好的，会破坏事件的因果性。彭罗斯针对这一困难提出宇宙监督假设：

“存在一位宇宙监督，它禁止裸奇异的出现。”

提出这一假设，是和西方的文化传统有关的。欧洲的文化是继承和发展古希腊与古罗马的文明而来的。在古罗马的时候，城市里有一种监督官，他不许任何人穿衣服在街上走。

彭罗斯觉得，为了保证时空中的因果性，宇宙也应该有一个“监督官”，他不许奇点和奇环不穿衣服裸露出来，这衣服就是黑洞的“视界”。由于视界的存在，黑洞外的人得不到任何来自黑洞内部的信息，当然也就看不见洞内的奇点和奇环，不会受它们的“不良影响”，因果性也就不会被破坏。

但是，这位宇宙监督是谁，彭罗斯没有说。这句话有点像大气压发现之前，有人提出的“自然害怕真空”一样，实际上没有给出问题的答案。

“宇宙监督”当然应该是一条自然定律，它可能是我们目前还不知道的什么定律，然而，自然界的基本定律不会太多，所以它更可能是一条我们已经知道的重要定律，只不过还没有把这条定律与“裸奇异”有机地联系起来。

从彭罗斯图5-4不难看出，进入克尔-纽曼黑洞，到达III区或者II'区的观测者，可以收到来自 $r=0$ 处（类时奇点或类时奇环）的信息。为了

防止进入黑洞的观测者看到奇点（以后把奇点和奇环统称为奇点），受到奇点的不良影响，彭罗斯又改进自己的“宇宙监督假设”，把它重新表述为

“类时奇异性是不稳定的。”

这是因为，研究表明，只要稍加扰动（例如有质点或飞船进入黑洞并抵达内视界附近）， $r=0$ 处的类时奇点就会倒下来，塌到内视界上，成为类光或类空奇点，阻断物体进入III区的通道（图5-8）。感兴趣的读者可以参看有关的广义相对论书籍。

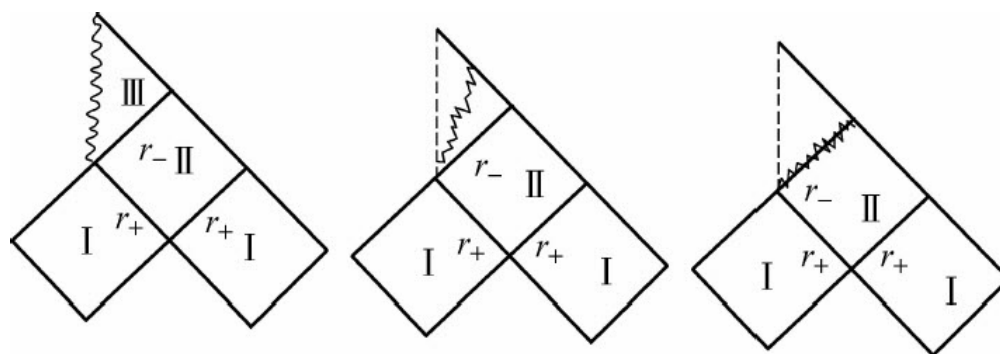


图5-8 不稳定的类时奇异性

无毛定理

研究认为，黑洞的边界是它的事件视界（简称为视界），外部观测者只能探知视界之外的事物，视界以内的任何物质和信息都跑不出来，外部观测者当然也就无法了解。

史瓦西黑洞的视界由它的总质量 M 决定。 $R-N$ 黑洞的视界由它的总质量 M 和总电荷 Q 决定；克尔黑洞的视界由它的总质量 M 和总角动量 J 决定；克尔-纽曼黑洞的视界则由它的总质量 M 、总电荷 Q 和总角动量 J 共同决定。

研究表明，最一般的稳态黑洞（即不随时间变化的黑洞）就是克尔-纽曼黑洞。对于外部观测者，最多只能探知稳态黑洞的三个信息：总质量 M 、总电荷 Q 和总角动量 J ，其他任何信息都无从了解。

人们把黑洞的信息戏称为“毛”，提出一个无毛定理：对于外部观测者，黑洞失去了除总质量、总电荷、总角动量之外的全部信息。

黑洞是“无毛”的。实际上还有三根毛，即 M 、 Q 、 J 。如果是我们中国人提出这一定理，大概就要戏称为“三毛定理”了。

无毛定理告诉我们，黑洞是一颗“忘了本”的星，它忘记了自己的过去，忘记了自己的历史，忘记了自己是由什么化学成分、什么物质结构的材料，在什么时候，在什么条件下形成的。

对于外部观测者，几乎失去了形成黑洞的物质的全部信息。不过，信息并没有从宇宙中消失，它们只不过被锁在了黑洞的内部。

彭罗斯过程

数学家彭罗斯在物理学家斯亚玛的动员下，进入了相对论研究的领域。他对黑洞的奇点和能层都进行了研究并做出了重要贡献。

首先，他提出并证明了奇点定理：认为符合广义相对论的物理时空一定至少存在一个内禀奇点，即时空曲率是发散的，不能通过坐标变换消除的本性奇点。

彭罗斯还创造性地把时空奇点看作时间开始和终结的地方，认为一定存在时间有开始或结束的物理过程。正是这一革命性的思想，把霍金吸引到黑洞的研究领域。对于这方面的情况，本书下面会详细介绍。

现在我们来看彭罗斯的另一研究，他提出一种物理过程，可以从转动黑洞的能层中提取能量，这就是著名的彭罗斯过程。在此之前，人们把各种黑洞都默认为恒星演化的最后归宿，都看作是死亡了的恒星。彭罗斯的结论告诉我们，至少转动的黑洞还存在进一步的演化过程。

彭罗斯发现，与黑洞外部的普通时空不同，转动黑洞的能层中存在负能轨道。一块能量为 E 的物质进入能层后，如果在适当位置分裂成两块，其中一块恰好进入负能轨道，并穿过事件视界落入黑洞，能量为 E_1 。另一块飞出黑洞，能量为 E_2 。由于 E_1 为负，能量守恒定律告诉我们，必有 $E_2 > E$ 。这等于说出射物体的能量大于了入射物体的能量，有黑洞的能量从这一过程中被提取出来。这就是彭罗斯过程（图5-9）。

研究表明，彭罗斯过程提取的能量是黑洞的转动能量，这部分能量原本储存在转动黑洞的能层中。在彭罗斯过程中伴随着能量提取出来的

还有黑洞的角动量。

这一过程使黑洞的转动逐渐变慢，能层逐渐变薄，最后转动停止，能层消失，克尔黑洞蜕变成史瓦西黑洞。如果是克尔-纽曼黑洞，则会退化成为R-N黑洞，即带电史瓦西黑洞。

当然，如果黑洞和入射物体均带电，这一过程也有可能提取黑洞所带的电荷和电磁能。这时黑洞将向史瓦西黑洞蜕化。不过，彭罗斯过程的前提是黑洞要存在能层。

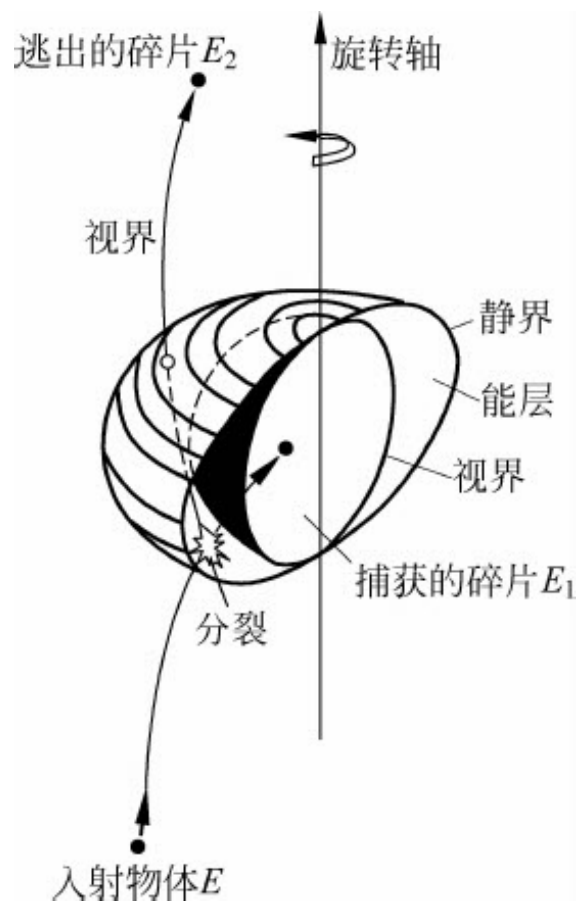


图5-9 彭罗斯过程

米斯纳超辐射

美国相对论专家米斯纳发展了彭罗斯的这一成果。他想到，如果入射物体很小，小到量子效应不能忽略，入射物体不就成了入射波吗？入射波是否也可以用来提取黑洞能量呢？他深入研究后果然发现，在一定的能量范围内，的确可以用入射到能层的波提取出黑洞的转动动能，也就是说，会有出射波能量大于入射波能量的现象出现，这一效应被称为米斯纳超辐射。形成超辐射的条件是入射粒子的静止质量 μ 要满足

$$m\Omega_H + eV > \mu \quad (5.25)$$

式中， Ω_H 和 V 分别是黑洞表面（即事件视界）的转动角速度及黑洞两极处的静电势， e 为粒子所带电荷， m 为入射粒子角动量在黑洞转动轴上的投影（类似于量子力学中的磁量子数）。

米斯纳超辐射是彭罗斯过程的量子表现。它提取的也是黑洞的转动动能和角动量。当然，如果粒子与黑洞都带电，也可以提取黑洞所带的电荷与电磁能。

对自发辐射的猜想

米斯纳对彭罗斯过程的发展，启发了苏联物理学家斯坦诺宾斯基和加拿大物理学家安鲁。他们想到，根据爱因斯坦的辐射理论，物质的辐射有两种，一种是自发辐射，另一种是受激辐射，这两种辐射的辐射系数之间有关联，存在受激辐射的物质，必定同时存在自发辐射。

超辐射本质上就是一种受激辐射。这样看来，转动或带电的黑洞似乎还应存在自发辐射，他们很快证实了自己的猜测。他们的研究结果被称为斯坦诺宾斯基-安鲁效应。

我们在介绍这一效应之前，需要先简单讲一下爱因斯坦的辐射理论。然后再简介一下相对论量子力学中的负能困难，以及有趣的狄拉克真空与反物质的概念。

爱因斯坦的辐射思想

爱因斯坦考虑原子中的两个能级， E_k 和 E'_k 。如果有能量恰为二能级能量差的光量子

$$h\nu_{k'k} = E'_k - E_k \quad (5.26)$$

射入，则处于低能级 E_k 的原子将吸收此辐射跃迁到高能级 E'_k ，这就是原子的吸收效应。如果有处于高能级 E'_k 的原子跃迁到低能级 E_k ，则会有 $h\nu_{k'k}$ 的光量子射出，这就叫原子的辐射。

但是，爱因斯坦指出原子的辐射有两种，一种是自发辐射，另一种是受激辐射。所谓自发辐射，就是在没有外界干扰的情况下，处于高能级的原子会以一定的概率自发地跃迁到低能级，并放出能量为 $h\nu_{k'k}$ 的光量子，如图5-10所示。

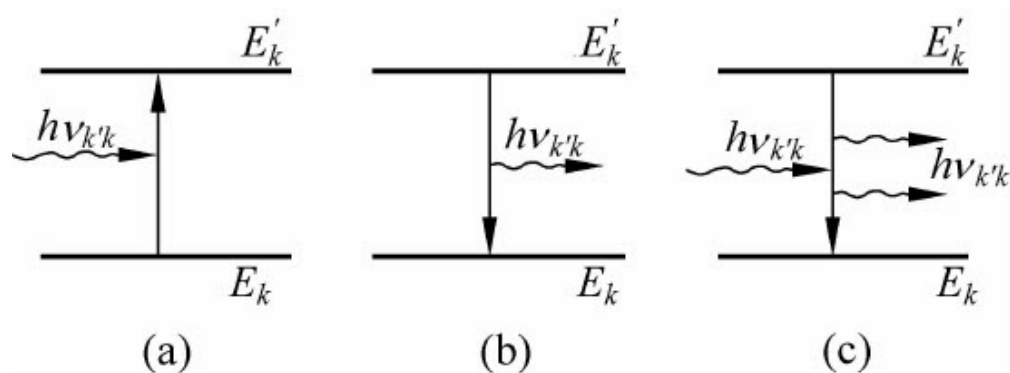


图5-10 爱因斯坦的辐射理论

(a) 吸收； (b) 自发辐射； (c) 受激辐射

这是学过物理的人早就知道的，不过爱因斯坦指出还存在另一种辐射情况，当有能量恰为二能级能量差的光量子 $h\nu_{k'k}$ 经过原子附近时，会

刺激处于高能级的原子，使它们一下子都跃迁到低能级，并放出同频率因而也同能量的光辐射，这就是所谓的受激辐射。提出“受激辐射”这一概念，是爱因斯坦的又一创举。

今天的激光理论和激光器，就建立在爱因斯坦提出的受激辐射理论的基础上。不过当年的爱因斯坦没有想到受激辐射会有这么大的应用价值。

激光的特点

由于受激辐射产生的光量子，与入射光量子同频率、同偏振、同相位、同方向，而且可以形成强脉冲，这就使激光器能够获得强大且单色性、平行性均非常好的光能，因而具有广泛的用途。

为了形成很强的光脉冲，可以使用光 and 电等方法，把处于低能级的原子大量激发到高能级，形成处于高能级的原子远远多于低能级原子的状态，即所谓“粒子数反转”的状态。这时如果有一个能量为 $h\nu_{k'k}$ 的光子射入，处于高能级的原子就会同时产生受激辐射，一瞬间像雪崩一样射出大量同频率、同偏振、同相位、同方向的光子，这就是著名的激光。

激光的特点是能量集中，首先是能量在频率上集中、且偏振方向一致、相位一致，因而单色性好、相干性好；其次是能量在空间上，也即在方向上集中，平行性好；第三是能量在时间上集中，可形成强脉冲。所以激光具有强大的功率和良好的相干性，因而在科研、军事和工业上有重要用途。

为了说明转动、带电黑洞附近粒子的自发辐射，我们先介绍一下有趣的狄拉克真空与反物质。

负能困难

我们现在讲一下相对论性量子力学中的负能困难。大家都知道，一个粒子的相对论能量 E 和动量 p 的关系是

$$E^2 = m_0^2 c^4 + p^2 c^2 \quad (5.27)$$

式中 m_0 是粒子的静止质量。上式不同于非相对论粒子的能量动量关系

$$E = \frac{p^2}{2m} + V \quad (5.28)$$

式中 V 是势能。把式（5.27）开方，可得粒子能量为

$$E = \pm \sqrt{m_0^2 c^4 + p^2 c^2} \quad (5.29)$$

我们看到，在相对论情况下，不但有正能粒子，而且有负能粒子。可是谁见过负能粒子呢？谁也没有见过。而且，粒子的负能量可以非常负，能量似乎没有下限。按照式（5.29）可以画出相对论性粒子的能级图。

从图5-11可以看出，粒子的负能量没有下限。从图中还可以看出，存在能量最低的正能粒子态，那就是动量 p 为零的静止正能粒子， $E = m_0 c^2$ 。同时，还存在最高的负能粒子态，那就是 p 为零的静止负能粒子， $E = -m_0 c^2$ 。在 $E = \pm m_0 c^2$ 之间是禁区，不存在能量介于 $+m_0 c^2$ 和 $-m_0 c^2$ 之间的粒子。

这一能级分布图告诉我们，正能粒子是不稳定的，因为量子效应使

得它可以通过跃迁飞过禁区进入负能态，同时放出能量，例如辐射出光量子。这一过程类似于原子中的能级跃迁（轨道间的跃迁）。

在玻尔原子模型中，电子只能存在于各个量子化的轨道（能级）上，各轨道之间的区域（即各能级之间的能态）电子是不能占据的，类似于图5-11中的禁区。然而，电子可以从高能轨道跳过“禁区”跃迁到低能轨道，同时射出光量子。

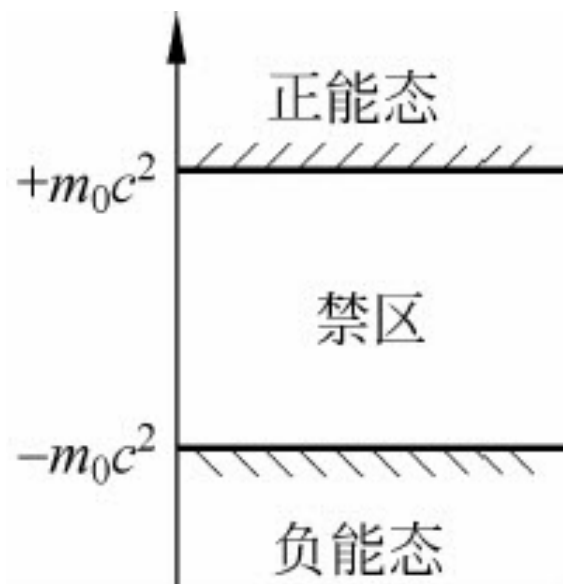


图5-11 相对论性粒子的正负能级

上面的能级图告诉我们，正能粒子可以跳过禁区跃入负能态，而处于负能态的粒子还可以进入更负的能态，同时放出能量。这一过程可以没有止境，粒子可以不断地、无限制地往更低的负能态跑，同时无限制地放出正能量。可是这一现象谁也没有看到过，这是怎么回事呢？

狄拉克真空

狄拉克提出了一个革命性的思想：真空不空。他认为，什么是真空呢？真空就是能量最低的状态。一个正能粒子都没有的状态是不是就是能量最低的状态呢，他认为不是。

可以打一个比方，一分钱都没有的人还不是最穷的人，因为还有欠债的人呢。欠债越多的人越穷，如果一个人一分钱都没有，并且欠债欠到最多，他把可以借钱的所有地方都借遍了，而且都借到了可能借到的最大值，他才是一个最穷的人。

同样，对于作为能量最低状态的真空，不仅应该一个正能粒子没有，而且应该所有的负能态都填满。所以，狄拉克认为，真空态就是正能态全部空着，而负能态全部填满的状态。

他认为不同粒子有不同的真空，例如电子真空、质子真空，等等。他以电子真空为例来说明自己的真空思想。

我们把图5-11看成电子的能级图。电子真空就是所有正能态都空着（一个正能电子都不存在），而且所有负能态都被电子填满的状态。泡利不相容原理要求一个状态只能存在一个电子，由于负能态都已填满，所以即使存在正能电子，这个正能电子也就只能永远处于正能态，不可能跃过禁区进入负能态，当然也不存在负能粒子落向更负能态的可能性。所以，狄拉克真空是稳定的，也是不空的。

打击真空

怎么证明狄拉克的真空图像是正确的呢？狄拉克说，我们可以打击真空，通过打击，赋予处于负能态的电子能量，使它们跃迁到正能态（图5-12）。这样就可以从真空中打出正能电子，同时留下一个负能空穴。由于电荷守恒，新生的正能电子带负电，所以负能空穴应该带正电。

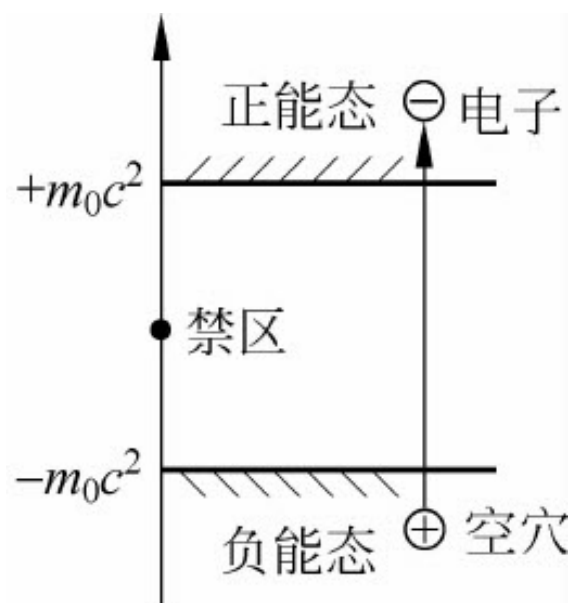


图5-12 打击真空

狄拉克最初猜测“负能空穴”也许是质子？因为当时学术界知道的带正电的粒子主要是质子。后来，很快弄清楚了，负能空穴不可能是质子。

因为把一个具有 $E = -m_0c^2$ 的电子打到正能态，变成 $E = m_0c^2$ 的正能电子，需要使这个电子跃迁过禁区，而禁区宽度是 $2m_0c^2$ ，所以外界必须提供给这个跃迁电子 $2m_0c^2$ 的能量。但跃迁完成后，生成的电子只具有 m_0c^2 的能量，另一部分 m_0c^2 的能量哪里去了呢？能量必须守恒，由于真空的其他部分都没有变化，只是多了个“负能空穴”，自然的想法是，“负能空穴”具有了 m_0c^2 的能量。所以，负能空穴是一个正能的粒子，由于它的质量与电子相同，但带正电荷，所以被称为“正电子”。于是狄拉克预言了正电子的存在。

狄拉克说，如果我的理论是正确的，应该能够从真空中打出正负电

子对。他认为，可以把原子核看作铁砧，重粒子看作铁锤，真空就夹在“铁砧”和“铁锤”之间。用铁锤（重粒子）打击夹在锤和砧之间的真空，即用重粒子轰击原子核，就可以从夹在二者之间的真空中打出正负电子对（即正能电子和具有正能的“负能空穴”）。

赵忠尧的贡献

1929年，中国旅美物理学家赵忠尧首先在实验中观测到了正负电子对的产生。1930年，他又观测到了正负电子对的湮灭。赵忠尧先生对正电子的发现作出了重大贡献。

遗憾的是，当时他头脑中没有“正电子”的概念。他把正电子的轨迹当成了反向运动的电子的轨迹。而且，当时还出现了几个与他的工作相矛盾的实验，那些实验似乎表明赵先生的实验有误。

然而，事实恰好相反。后来的实验表明，那些不利于赵先生的实验都是错的，只有赵先生的实验是正确的。遗憾的是，时机已经错过，本该获得诺贝尔奖的赵忠尧先生与这一大奖失之交臂，发现正电子的诺贝尔奖不恰当地只给了安德逊一个人。

安德逊在研究宇宙线轰击原子核时，打出了正负电子对。当时认为，这就是作为“铁锤”的宇宙线粒子，打在作为“铁砧”的原子核上，从夹在“锤”和“砧”之间的真空中打出了电子对。不过，进一步的研究表明，正负电子对并非宇宙线直接轰击原子核得到的，而是宇宙线在穿过空气时，

撞在空气中的粒子上产生了 γ 射线， γ 射线再撞在原子核上，在这一撞击过程中 γ 射线（即光子）生成了正负电子对（图5-13）。

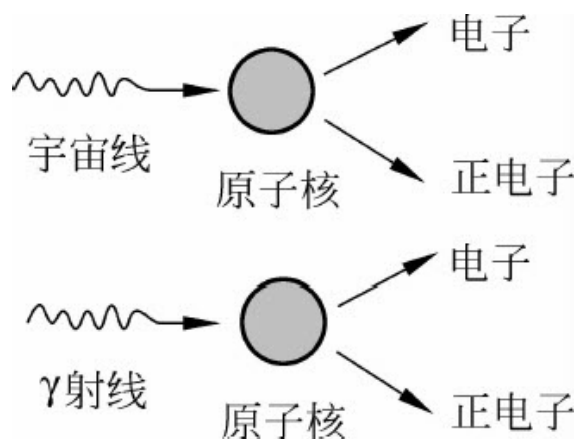


图5-13 正负电子对的产生

反物质

正电子发现之后，人们又发现了反质子，即“负能质子的空穴”。反质子与质子质量相同，但是带负电荷。反质子与正电子一起，可以构成“反氢”原子。“反氢”是反物质的一种。

所谓反物质，其原子核带负电，由反质子和反中子组成（反中子与中子相似，但是磁矩相反），正电子围绕着“反核”转动。值得注意的是，反物质和通常的物质一样，质量和能量都是正的。

王淦昌的贡献

我们中国人对基本粒子研究作出的另一个大贡献是王淦昌先生发现了反西格玛负超子。他是在苏联杜布纳联合核子研究所作出这一发现的，这个研究所是苏联、中国等12个社会主义国家合作组建的。

观察到反西格玛负超子的丙烷气泡室是王淦昌先生设计的，这一设计是完成此项发现的关键一步。

杜布纳联合核子研究所只发现了这一种基本粒子，王淦昌先生功不可没。

赵忠尧和王淦昌先生早年就放弃了国外优越的实验条件和优裕的生活条件，回到贫穷、落后的祖国，为中国核物理和粒子物理的发展做出了重大贡献，培养了大批人才。

转动、带电黑洞视界附近的狄拉克能级

研究表明，转动和带电的黑洞，会使视界附近的狄拉克能级发生变化，如图5-14所示。

从图5-14可以看出，狄拉克能级在这类黑洞的表面（视界）附近会“提升”，其中一部分负能级（图中交叉线表示的阴影部分）的高度会高于邻近的正能级。按照狄拉克真空的思想，这部分负能级是充满电子的，而禁区右上方的正能级是空着的，所以阴影部分右方的禁区好像是一个势垒，研究表明，阴影区的电子会通过隧道效应穿过这一“势垒”，到达右方的区域成为正能电子。

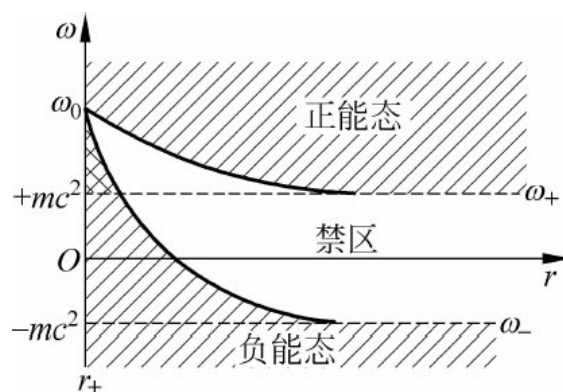


图5-14 转动、带电黑洞视界附近的狄拉克能级

这就是黑洞的自发辐射，它会带出黑洞的转动能、角动量、电磁能和电荷。使转动带电的黑洞蜕化为不转动、不带电的史瓦西黑洞。这时，黑洞附近的狄拉克真空能级将蜕化成图5-15的样子。

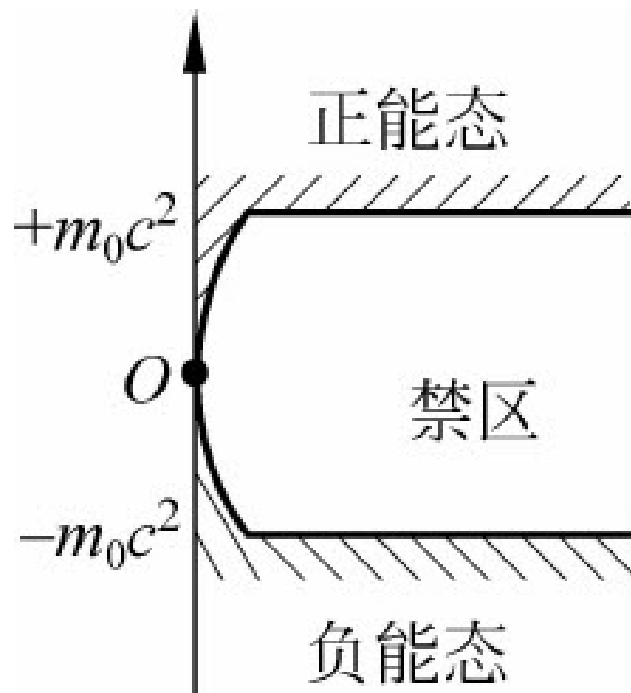


图5-15 史瓦西黑洞视界附近的狄拉克能级

我们看到，转动和带电的黑洞还是有生命力的。它们存在彭罗斯过程，以及自发辐射和超辐射。转动带电的黑洞好像是黑洞的激发态，史瓦西黑洞则好像是黑洞的基态。当各种黑洞蜕化为史瓦西黑洞之后，黑洞似乎终于真的成了死亡的星，成了恒星的最后归宿。



绘画: 张京

第六章 探索黑洞的明星——霍金

现在来介绍黑洞研究中的一颗明星——霍金（图6-1）。

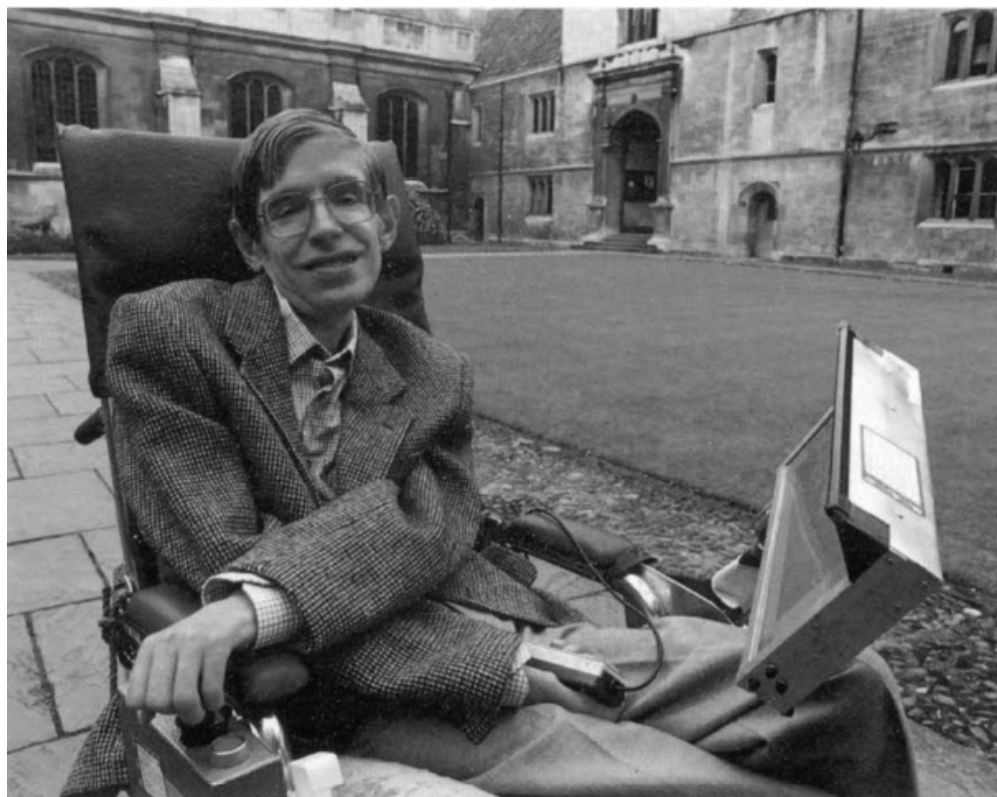


图6-1 霍金在剑桥大学校园

少年霍金

霍金1942年1月8日出生于英国的牛津，这一天正是伽利略逝世300周年的纪念日。霍金经常跟人提起：“我是在伽利略逝世300周年的那一天出生的。”似乎在提醒人们，你们看我像不像伽利略再世？不过，他同时也说，其实在那一天出生的孩子有20万。

他出生在牛津，并非他的家在牛津。当时“二战”正在进行，为了不损毁英国的文化中心，德国飞机不炸牛津和剑桥，作为回报，英国飞机也不炸德国的哥丁根和海德堡。

霍金的父母都毕业于牛津大学。父亲搞生物医学。他出身贫寒，很注意节俭，但对别人很慷慨。母亲学文科，长期做文秘工作。她年轻时是英国共青团的团员，后来转入了比较温和的工党。她经常带着年幼的霍金参加群众集会和游行，对少年霍金产生了较大影响。

由于家境不富裕，上不起私立的贵族子弟学校，少年霍金上的是普通中学，还曾在女子学校上过学，那所学校接收幼年的男生。

霍金上学时正在搞教改，把学校分成学术性和非学术性两种，在学术性学校中又把学生按成绩好坏分在不同的班。他所在的年级分为A、B、C三个班。成绩最好的在A班，中等的在B班，差的在C班。每学年三个班的学生进行调整。A班20名之后的学生要降到B班，B班的前20名升到A班。B班和C班也做类似的调整。

霍金第一学期考了第24名，第二学期考了第23名，幸亏有个第三学期，他考了第18名，才没有降到B班。霍金本人对这种教学方式很不赞同，认为对那些降下去的同学打击太大了，自信心受到很大损害。老子

《道德经》中有一句话“大器晚成”，在中国古代颇有影响。这种教学方式显然不利于大器晚成的人。

霍金不是班上最拔尖的学生，同学们感觉他的实际水平在第10名左右。不过，他经常看课外书，知识面较宽，爱独立思考，喜欢与同学们交流。诸如“宇宙的起源需不需要上帝帮忙啊？”“远方星系的红移是否由于光线在旅程中走得太累了，从而变红啊。”

他成绩一般，字也写得不好，老师并不怎么看好他，但似乎同学们看出了什么苗头，给他起了个外号叫“爱因斯坦”。结果这个绰号叫发了，他后来真的被誉为“当代的爱因斯坦”。

中学时他并不喜欢物理课，觉得简单而枯燥。他认为化学就有意思多了，经常会发生一些意想不到的事情，例如爆炸啦，起火啦，等等。

考大学时，父母希望他学医，将来当医生。医生救死扶伤，是个受人尊敬的职业，而且在西方收入也不错。但是他在高中时受到一位老师的影响，对物理产生了兴趣。觉得物理研究的问题，小至基本粒子，大至整个宇宙，有趣而高深，让人觉得神秘莫测。于是他报考了牛津大学的物理系。由于紧张，考前有点失眠，但发榜时还是考上了。

在牛津的大学生涯

当时英国的大学也在搞教改。学生只在刚入学时考一次试，然后就不考了，只在毕业前再考一次，用4天时间上下午连续考，把三年学习的内容统考一遍。由于学校的平时管理放得很松，学生们相当自由。霍金喜欢玩，尤其喜欢划船，结果用到学习上的时间就很少，霍金后来回忆，觉得自己大学期间每天的学习时间平均只有一小时。

有一次上电磁学课，老师布置大家看课本的某一章，并把这章后面的13道题都做了，然后就宣布下课，两个星期后再见。下课后，霍金一直在玩，直到两周后交作业的前一天，才想起自己还没有做作业。那天早晨，同学们又约他去玩，他说不行了，必须赶紧补作业。那几个同学暗想：你今天才想起做作业，你知道这题有多难吗？你试试吧。他们几个人努力了两个星期，每人才做出一两道题。中午，这些同学返回宿舍时，正好碰到霍金下楼去吃饭，于是问他：“题做得怎么样了？”他说：“这题真难，我没来得及做完，只做了10道。”

疾病来袭——人生的转折

霍金人生的转折发生在大学毕业的前一年。他有一次系鞋带时突然发现手不灵活了。英国的医学很先进，很快确诊他患了进行性肌肉萎缩。这是一种不治之症，肌肉萎缩会不断加重，人逐渐走向全身瘫痪，医生预言他最多还能活两年。“哎哟！年轻人怎么得了这种病？吃点好的吧。”

霍金心情十分沮丧，自己的青春才刚刚开始，没想到生命马上就要结束了。他买了一些啤酒，在宿舍中一个人独饮。最初他什么事也干不下去，头脑中一片空白，这时他的女友安慰他，表示尽管他有病，自己还是要嫁给他，这给了霍金很大的鼓舞（图6-2）。

他想，反正一时半会儿还死不了，而且既然要结婚，就要准备过日子，将来就要有职业。首先要有一个博士学位，所以自己还是应该努力，不然怎么求职？怎么养家？

于是他生平第一次开始努力学习。在勤奋的学习中他发现自己挺喜欢也挺适合科学研究的，而且在学习和研究中享受到很大的乐趣。他不禁想起了有人说过的一句俏皮话：“科学家和妓女都从自己喜爱的职业中获得了报酬。”



图6-2 霍金夫妇结婚时的照片

大学本科毕业时，同宿舍的4个人都报考了研究生。笔试结束后霍金和另外两个同学觉得沮丧，只有一个同学觉得考得不错。成绩公布时，除去那位觉得考得不错的同学感到失望之外，他们三人倒都比预想的要好。

口试时霍金已经拄着拐杖。考试委员会问他想去剑桥还是留在牛津。霍金回答说，你们要是成绩给我一等，我就去剑桥；要是给我二等，我就留在牛津。结果考试委员会给了他一等，于是他就去了剑桥。

霍金原本对粒子物理有兴趣，但那时的粒子物理重点是研究粒子分类和对称性，弱电统一理论和量子色动力学都还没有出现，他觉得这方面的研究有点与植物学分类类似，没有较为成熟的理论，看不出有什么前景，于是兴趣就转向了理论天体物理和宇宙学，因为那个领域有一个成熟的理论可用，那就是爱因斯坦的广义相对论。

霍伊尔与斯亚玛

霍金去剑桥是想追随著名的霍伊尔教授研究宇宙演化理论。我们在第二章中谈到，霍伊尔曾对恒星的聚变反应理论作出过重大贡献。这里我们要讲到他的另一重大成果，他是稳恒态宇宙模型的创始人。稳恒态宇宙模型与大爆炸模型不同，认为不存在原始火球。宇宙在膨胀过程中不断有物质从真空中产生，宇宙演化中物质的密度和温度都没有大的变化。

大爆炸宇宙模型原本称为火球模型。这一模型认为宇宙起源于一个原始的核火球，物质从核火球中创生出来，在宇宙膨胀过程中不再有物质从真空中产生，宇宙中物质的密度和温度都随着膨胀而减小。火球模型是由俄裔物理学家伽莫夫提出的。

霍伊尔坚决反对这一模型，认为这个原始火球无异于一场大爆炸，于是讥讽它为大爆炸模型，后来“大爆炸模型”这个称号就流传下来了。

霍金原想追随霍伊尔学习，没想到霍伊尔不要他。当时剑桥还有一位研究广义相对论和宇宙学的老师，叫斯亚玛。斯亚玛是谁？霍金从来没有听说过。霍伊尔不要他，霍金没有办法，只好去找斯亚玛，斯亚玛接收了他。

斯亚玛以不主动管学生著称。你要来当我的学生，可以，但我不给你找科研题目，你自己找。如果你不来找我，我不会主动去管你，你毕不了业自己负责。你如果来找我，问我问题，我可以跟你讨论，可能建议你去找某个人，或去看某本书、某篇论文。

笔者听说斯亚玛这种带博士生的方式时，觉得这个导师太不负责任

了，简直不合格啊。后来了解到，当时世界上的几个最优秀的青年相对论天体物理学家中，几乎有一半出自斯亚玛的门下，笔者才意识到自己最初看法的错误。

博士生确实应该像斯亚玛那样带，对待博士生不应该像硕士生那样。因为硕士生一般没有参加过科研，老师给他题目，领他入门，并帮助他前进是可以的，让他从研究中学会研究。而博士是高级人才，应该具有自主科研的能力，应该学会自己找题目，自己攻关，老师最多辅助一下，这样才能成长为合格的博士。

斯亚玛还有一个优点，经常在自己办公室待着，学生很容易找到他。霍伊尔则不同，整天在外边奔走，学生很难找到他。

初露锋芒：与霍伊尔的争论

一开始，霍金没有课题，脑子里还在想着霍伊尔的研究领域。他就经常到霍伊尔的研究生纳里卡的办公室去。国外的研究生都是一个人一间或两个人一间办公室，里面每人一张桌子，一把椅子，一个书架。办公室里通常还有一块黑板，一个电话。纳里卡是一个人一间办公室。

霍金进去后看见纳里卡正在计算着什么。就问他，你在算什么呢？原来是搞霍伊尔建议他做的题目。霍金说：我帮你算好吗？纳里卡说，当然好了。于是霍金就帮他计算。在研究过程中，霍金发现霍伊尔的新理论中有个致命错误：其中的一个系数实际上是无穷大！大家知道，系数只能是有限值，既不能是零，也不能是无穷大。霍金反复论证后，觉得这个系数肯定是无穷大。

不久，霍伊尔在伦敦的一次学术会议上报告他的这一工作，他还不知道霍金已经发现了他的新理论的错误。做完报告后，霍伊尔严肃地扫视了一下会场上的100多位听众，问：“有没有问题？”

这时坐在最后一排的霍金拄着拐杖站了起来，说：“我有一个问题。”霍伊尔问：“什么问题？”霍金说：“你报告中的那个系数是无穷大。”霍伊尔说：“不是无穷大。”霍金说：“是无穷大。”听众中传出一阵令霍伊尔难堪的笑声。霍伊尔明白，如果霍金说的问题确实存在，自己的工作就全错了。于是他板起面孔问：“你怎么知道？”霍金说：“我论证过。”会场上又出现笑声，霍伊尔恼羞成怒。

霍伊尔认为霍金是不道德的，既然发现了论文的论文的错误，为什么不事前在会下提出，要到会上来让我出丑。霍金和其他一些年轻人则认

为，真正不道德的是霍伊尔教授本人，为什么把不成熟的工作拿到大会上做报告？霍金挑出的这个重要失误，使霍伊尔企图改进稳恒态理论的新模型泡了汤。

那段时期的天文观测对稳恒态理论不利，特别是发现了3K微波背景辐射，科学家们很快确认那就是一直在寻找的宇宙大爆炸的余热。这一发现极大地支持了大爆炸模型，稳恒态宇宙模型遭到重大挫折，从此，大爆炸理论成为宇宙演化的主流理论。

幸会彭罗斯

斯亚玛对霍金的研究能力非常赞赏，他介绍霍金认识了数学家彭罗斯。彭罗斯是在斯亚玛的劝说下进入相对论研究领域的，这时他刚刚做了一件重要工作，提出了“奇点定理”。彭罗斯证明，按照广义相对论，在弯曲时空（即万有引力）作用下的塌缩星体，不仅有可能形成黑洞，而且形成黑洞后所有的物质都要会聚到黑洞的中心，在那里形成物质密度和时空曲率为无穷大的奇点。

彭罗斯创造性地对奇点作了重新认识，认为奇点就是时间开始或结束的地方，黑洞内部有个时间的终点，白洞内部有个时间的起点。彭罗斯的奇点定理说，只要时空的因果性良好，广义相对论正确，而且有一点物质，黑洞情况的时间演化就一定有一个终结，也就是说黑洞内部一定会存在时间的终点——奇点。

霍金原本不熟悉彭罗斯使用的现代微分几何（即整体微分几何），但他对彭罗斯的奇点定理大感兴趣，而且猜想到彭罗斯的结论也许可以适用于宇宙演化。于是他开始学习使用这种几何，并努力把它应用于宇宙演化。

经过努力霍金终于证明大爆炸宇宙必定有一个初始奇点（时间开始的时刻），大塌缩宇宙则必定有一个终结奇点（时间结束的时刻），把彭罗斯的奇点定理推广到了宇宙学情况。不过，他的第一篇论文后来查出有缺陷，发现后，终于进一步给出了正确而严格的证明。

霍金后来的博士论文的前一半，就是论述霍伊尔模型的错误。后一半则是发展彭罗斯的奇点定理。他如愿以偿地获得了博士学位，他的科

研能力也得到斯亚玛和彭罗斯等专家的高度评价。

面积定理

1970年的一天，霍金正要上床睡觉，突然想到他和彭罗斯建立的证明奇点定理的方法，有可能用于研究黑洞的变化和碰撞，他立刻认识到黑洞的表面积 A 随着时间的发展，只能增大，不能减少。这就是所谓的“面积定理”：

$$dA \geq 0 \quad (6.1)$$

按照这个定理，两个黑洞可以合并成一个，但一个黑洞绝不可能分裂成两个。而且，通过两个黑洞合并前后的质量变化，可以计算出黑洞碰撞能放出多大能量。当时霍金的心情非常激动，简直彻夜难眠。第二天他仔细地完成了这篇研究论文。

贝肯斯坦的大胆创新：黑洞有熵和热

霍金的面积定理引起了美国一位年轻的研究生贝肯斯坦的极大兴趣。贝肯斯坦想，怎么会出现这样一条定理呢？黑洞的表面积为什么只能增加不能减少呢？太奇怪了。他突然想到热力学中的熵，按照热力学第二定律，孤立系统（或绝热系统）中的熵只能增加不能减少。黑洞的表面积与熵有点类似，黑洞的表面积会不会就是黑洞的熵呢？

贝肯斯坦把自己的想法告诉了老师惠勒教授。惠勒是当时世界上最优秀的相对论专家，他曾经参加过美国的氢弹研制，后来转入相对论研究。惠勒非常欣赏贝肯斯坦的猜想，自己的学生也许是对的，这可是个重要发现。于是他和贝肯斯坦抓住这一问题进行了讨论。

在讨论中，惠勒讲，假如你手中拿着一杯热水，它有温度也有熵。如果这时恰好有一个黑洞从你身边飘过，你把这杯水扔进黑洞，由于我们不能从黑洞中提取任何物质信息，那么这杯热水连同它携带的熵都从我们身边消失了，自然界中的熵不就减少了吗？这可是违背热力学第二定律的啊！

贝肯斯坦答道，这杯水扔进黑洞后，黑洞的质量增加，它的表面积也随之增加，热水的熵虽然消失了，黑洞的面积熵却增加了。这个思想实验正好支持了黑洞表面积本质上是“黑洞熵”的猜想，如果这一猜想正确，热力学第二定律就不会被破坏了。他们非常高兴，看来黑洞的表面积确实是熵。

贝肯斯坦-斯马尔公式

在热力学中，熵与温度是成对的物理量，黑洞既然有熵，似乎也应该有温度。黑洞真的有温度吗？答案是肯定的。

贝肯斯坦和另一位学者斯马尔在研究黑洞的几个参量的函数关系时，分别得到了一个积分公式

$$M = \frac{\kappa}{4\pi}A + 2\Omega J + VQ \quad (6.2)$$

和一个微分公式

$$dM = \frac{\kappa}{8\pi}dA + \Omega dJ + VdQ \quad (6.3)$$

式中， A 为黑洞表面积； κ 为黑洞表面引力，粗略地说它就是单位质量的质点静置于黑洞表面（视界）上时所受的引力； Ω 、 J 分别为黑洞视界的转动角速度与黑洞的角动量； V 、 Q 分别为黑洞两极处的静电势和黑洞所带的电荷； M 为黑洞质量，由于采用了自然单位制， $c=1$ ， M 也就是黑洞的能量。

积分公式又称斯马尔公式，微分公式又称贝肯斯坦公式。把积分公式两端微分就可得到微分公式，读者不妨试一下。不过要注意，式（6.2）中的 M 、 κ 、 A 、 Ω 、 J 、 V 、 Q 均是变量，要同时对这7个量微分，才能从式（6.2）导出式（6.3）。

贝肯斯坦把微分公式和热力学第一定律比较，

$$dU = TdS - pdV \quad (6.4)$$

认为既然黑洞表面积 A 相当于熵，那么表面引力 κ 就相当于温度。式（6.4）中 U 、 T 、 S 、 p 、 V 即通常热力学系统的内能、温度、熵、压强和体积。式（6.3）右端第一项与式（6.4）第一项类似，为系统吸收的热量；式（6.3）的后两项与式（6.4）的后一项类似，都相当于系统对外所做的功。

霍金的反驳

黑洞是广义相对论预言的天体，面积定理是从广义相对论和微分几何推出的结论，这里面没有引入热效应和统计物理假定，黑洞表面积怎么可能是熵呢？黑洞怎么会出现热力学性质呢？如果黑洞有温度，就会有热辐射。但是广义相对论告诉我们，黑洞是只进不出的天体，内部的任何东西都跑不出来，怎么可能产生热辐射呢？霍金对贝肯斯坦等人的工作很不以为然，认为贝肯斯坦错误理解了自己的面积定理。

霍金和另外两位物理学家，巴丁和卡特，重新研究了贝肯斯坦等人的工作，认为式（6.2）与式（6.3）虽然正确，但并不表示热力学关系。式（6.3）像热力学第一定律但不是热力学定律；式（6.1）像热力学第二定律，但也不是热力学定律。霍金他们称其为黑洞力学第一定律和黑洞力学第二定律，认为它们只是力学定律，不是热力学定律。他们还类似得到了黑洞力学的第三定律和第零定律，如表6-1所示。

表6-1 黑洞力学定律与普通热力学定律的比较

	普通热力学	黑洞力学
第零定律	处于热平衡的物体，具有均匀温度T	稳态黑洞的表面上，κ是常数
第一定律	$dU = TdS + \Omega dJ + VdQ$	$dM = \frac{\kappa}{8\pi} dA + \Omega dJ + VdQ$
第二定律	$dS \geq 0$	$dA \geq 0$
第三定律	不能通过有限次操作，使T降到零	不能通过有限次操作，使κ降到零

霍金等人认为，黑洞表面积 A 像熵，但不是熵；表面引力 κ 像温度，但不是温度。黑洞没有温度，所以不会有热辐射，原来的黑洞理论仍然是正确的，黑洞就是一个只能进不能出的天体，洞内的任何物质都不可能跑出来。

霍金、巴丁和卡特是在参加一个暑期讨论班时完成上述研究的，在论文中他们明确表示了不同意贝肯斯坦等人对黑洞的热力学解释。

霍金观点的转变：黑洞有热辐射

讨论班结束之后，霍金返回剑桥。与贝肯斯坦的争论在他的头脑中无法消除，他在反对贝肯斯坦观点的同时，又觉得贝肯斯坦的怪异想法很有吸引力，很具启发性。他又忍不住回过头来想：万一贝肯斯坦是对的呢？如果贝肯斯坦是对的，黑洞真的有温度，就应该有热辐射，黑洞就一定会以黑体辐射的形式射出粒子。这可能吗？

经过反复思考研究，霍金把量子效应附加到黑洞上，用弯曲时空量子场论的方法证明，黑洞真的能辐射出粒子，粒子的能量分布是黑体辐射谱。黑洞表面引力 κ 恰恰是黑体辐射谱显示的温度。

所谓弯曲时空量子场论，就是引力场不量子化，仍看作经典场，但把物质场量子化，研究量子场方程在弯曲时空背景下的性质。采用这种半经典半量子的方法，是由于引力场量子化问题碰到了巨大困难，至今尚未解决。

霍金所用的弯曲时空量子场论方法，有点像玻尔的量子论，也有点像非相对论的量子力学，在这两种量子理论中，物质被量子化了，但电磁场没有量子化，仍看作经典场。历史表明，这样的半经典半量子化理论是能够解决不少问题的。

霍金的工作震动了理论物理和天体物理界，黑洞不再是只进不出的天体，考虑量子效应后它会产生热辐射，黑洞具有温度。为了纪念霍金的成就，黑洞的热辐射被称为霍金辐射。

在霍金做出这一成就后，他的老师斯亚玛评价说：“毫无疑问，霍金已成为20世纪最伟大的物理学家之一了。”斯亚玛又自豪地说：“我对

广义相对论有两个重要贡献，一是培养了霍金这个学生，二是把彭罗斯拉进了广义相对论的研究领域。”

黑洞里面是单向膜区，时间的方向指向奇点，怎么可能有粒子从黑洞中跑出来，形成热辐射呢？霍金在给出严格数学证明的同时，也给出了一个清晰的物理解释。有关讨论要从真空涨落谈起。

真空涨落

当时狄拉克的相对论性量子力学已经进一步发展成量子场论。狄拉克关于真空的思想也被进一步发展了。按照量子场论，真空并不是完全平静的，不断会有虚的正反粒子对产生，其中一个粒子是正能，另一个是负能，产生之后很快湮灭，这种现象叫真空涨落。

但是我们从来没有看到这种虚正反粒子对的产生，特别是谁也没有见过固有能量为负的粒子，这是为什么呢？这是由于测不准关系（即“不确定关系”）在起作用。根据测不准关系

$$\Delta t \Delta E \sim \frac{\hbar}{2} \quad (6.5)$$

由于虚粒子存在的时间 Δt 非常短，要想在 Δt 的时间内观测到它们，就会受到量级为

$$\Delta E \sim \frac{\hbar}{2\Delta t} \quad (6.6)$$

的能量干扰，这一能量恰好掩盖了虚粒子对的存在，所以我们不可能直接观察到虚粒子对，当然也就不可能观察到虚的负能粒子。虽然没有直接观察到真空涨落，真空涨落的间接效应却被普遍观察到了，所以真空涨落的观点和理论很快被学术界普遍接受。

霍金辐射

霍金利用真空涨落来解释黑洞辐射（图6-3）。黑洞外部的真空当然也会发生涨落。如果黑洞外部产生一对虚正反粒子，然后又湮灭，或者虚粒子对一起掉入了黑洞，这都不会产生特殊的效应，也不会导致黑洞辐射的产生。但是还有一种不对称的情

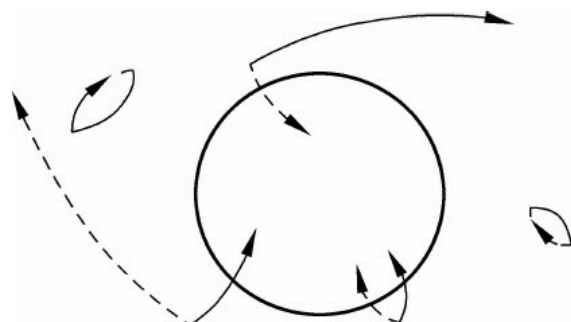


图6-3 黑洞附近的真空涨落

况，就是虚粒子对中负能的一个掉进黑洞，正能的飞向远方，这种情况就会导致黑洞辐射的产生。下面我们来详细解释一下。

假设在黑洞附近产生的虚粒子对是电子和正电子。电子是正物质，正电子属于反物质，通常提到的物质和反物质都是正能的，提到的电子和正电子当然也是正能的。不过我们现在讨论虚过程，在虚电子对中的电子和正电子，一个是正能，另一个是负能。可以是正能电子和负能正电子，也可以是负能电子和正能正电子。为了讨论方便，我们假定虚粒子对中电子带正能，正电子带负能。

现在负能粒子正电子落入了黑洞，正能电子飞向了远方。落入黑洞的负能正电子，顺着时间前进，穿过单向膜区到达奇点，与聚集在奇点处的物质汇合，使那里少了一个电子的质量（注意，电子和正电子只是电荷相反，质量和其他物理量都相同），多了一个正电荷。

远方的观测者看到一个电子从黑洞处飞过来了，它具有正能量和负电荷，同时看到黑洞少了一个电子质量，多了一个正电荷（注意，无毛定理还是允许黑洞存在三根毛的，所以外部观测者能够看到黑洞质量和

电荷的变化)。于是，他认为黑洞向他辐射了一个电子，这个电子带出来原本属于黑洞的一点点质量和电荷。

霍金又进一步阐述了这一物理图像。他说，视界（黑洞表面）附近真空涨落产生的虚电子对，其中的负能正电子落入了黑洞，顺着时间方向落向奇点，正能电子飞向远方；这一过程相当于黑洞奇点附近产生一个正能电子，它逆着时间穿过单向膜区到达视界，在那里被视界散射，再顺着时间进展飞向远方。

由于视界附近的真空涨落大量存在，会有大量负能粒子落入黑洞，与其成对的正能粒子大量飞向远方，这就构成了黑洞的辐射，霍金严格证明了这种辐射是黑体辐射，具有普朗克谱，也就是说，这是纯粹的热辐射，从辐射谱可以看出辐射温度，它也就是黑洞的温度。霍金证明这一温度恰为

$$T = \frac{\kappa}{2\pi K_B} \quad (6.7)$$

与贝肯斯坦给出的完全一致。不难看出，和上式对应的黑洞熵的表达式为

$$S = \frac{K_B A}{4} \quad (6.8)$$

式中， κ 为黑洞表面引力， A 为黑洞的表面积， K_B 为玻耳兹曼常数。

读者可能会产生一个疑问，真空涨落会不会产生相反的现象，正能粒子落入黑洞，而负能的一个飞向远方呢？科学研究回答说：不会。

这是因为普通的时空区（包括黑洞外部）不允许负能粒子单独存在，负能粒子只能在真空涨落过程中，在虚粒子对中瞬间出现，然后很

快湮灭。

黑洞内部的时空区则不同，它允许负能粒子长期存在。如果虚粒子对中正能的一个掉进了黑洞，洞外时空不允许负能粒子久留，所以负能的一个必定跟着掉进去。

因此不会出现“正能粒子落入黑洞，负能粒子飞向远方”的相反的效应。只会存在负能的掉进去，正能的飞向远方的效应，这一效应是不对称的。正是这一“不对称”，导致了黑洞热辐射的产生。

于是霍金证明了黑洞能产生热辐射，还真的有温度，当然也就真的有熵。为了纪念霍金的这一功绩，黑洞热辐射被称为霍金辐射。他一生最杰出的工作当推霍金辐射的发现。

霍金与贝肯斯坦的成就是历史性的。黑洞是广义相对论预言的天体，在论证黑洞存在时，人们只用了相对论、时空弯曲和几何知识，根本没有引进统计物理和热力学，却居然得到了黑洞有温度和熵的结论，这太奇妙了。

在证明黑洞辐射时用了量子理论的知识，但是只用了量子论的基本原理，没有涉及任何具体的物质结构模型。黑洞热性质似乎表明“热”与“万有引力”之间存在本质联系。

克尔-纽曼黑洞的温度

与史瓦西黑洞一样，克尔-纽曼黑洞也有温度，有热辐射。研究表明，这种黑洞的温度和熵也由式（6.7）和式（6.8）决定。

克尔-纽曼黑洞的表面引力的表达式比较复杂，由

$$\kappa = \frac{r_+ - r_-}{2(r_+^2 + a^2)} \quad (6.9)$$

表示。式中 r_+ 与 r_- 分别为黑洞外视界和内视界的位置，如式（5.10）所示。

克尔-纽曼黑洞的霍金辐射中含有自旋为 $\frac{1}{2}$ 的粒子（如电子、质子等），这一点是由我们课题组首先证明的。霍金证明球对称的史瓦西黑洞热辐射各种粒子后，很快就有人证明了克尔-纽曼黑洞热辐射自旋为整数的各种粒子（玻色子），但证明热辐射自旋为半整数的费米子却遇到了极大困难，关键是要用到弯曲时空中的旋量方程，这种方程中粒子波函数的各分量耦合得很厉害，很难求解。刘辽教授和许殿彦教授带领我们这个组，对有关方程做了退耦，并找到了严格解，最终完成了这一证明。

从式（6.9）可知，极端黑洞的 $\kappa=0$ ，也就是说，极端黑洞的温度是绝对零度。所以，有人推测“宇宙监督”就是热力学第三定律。第三定律认为：不可能通过有限次操作，把系统的温度降到绝对零度。对于黑洞来说，就是禁止黑洞演化成极端黑洞。极端黑洞尚存一张视界膜，如果达不到极端黑洞，当然就更不可能让这层膜消失，奇环也就裸露不了。

安鲁效应

在霍金提出黑洞有热辐射的前夕，安鲁（W. G. Unruh）发现，匀加速直线运动的伦德勒观测者处在热浴中。这就是说，原本一无所有的闵氏时空，所有惯性观测者均认为是真空，但是，在其中作匀加速直线运动的观测者会发现自己周围充满了热辐射，其温度为

$$T = \frac{a}{2\pi K_B} \quad (6.10)$$

这个温度取决于伦德勒参考系的加速度 a 。

安鲁的结论是惊人的。然而，由于大部分物理工作者不熟悉广义相对论，也由于这一效应过于微弱，目前在实验中还观测不到，这一杰出的工作至今还不为世人所注意，只有少数人知道有这个已被预言但尚未观测到的效应存在。

安鲁等人认为，伦德勒观测者感受到的热效应是一种量子效应，它是由于不同参考系有不同的“真空”而造成的。

通常的物理学都是在平直的闵氏时空的惯性系中讨论的，所以物理学中所说的真空，通常都是指惯性系中的真空。量子场论认为，不同惯性系中的真空是等价的，也就是说真空在洛伦兹变换下是不变的，因此以往没有发现“真空”会对参考系有依赖。

安鲁发现，越出惯性系去讨论“真空”，就会出现问题了。他发现加速参考系的真空和惯性系的真空有所不同，于是人们把惯性系中的真空称为闵氏真空，以区别于非惯性系中的真空。

按照狄拉克的思想，真空不空，有零点能存在。闵氏真空的虚粒子涨落形成零点能（图6-4）。此零点能对于闵氏时空中所有的惯性系都相同。然而，当我们在作匀加速直线运动的伦德勒系中观测时，由于伦德勒真空不是闵氏真空，它的能量零点比闵氏真空的能量零点要低，因此，闵氏真空的零点能在伦德勒观测者看来就是高于真空零点的能量，是真实可测的能量。

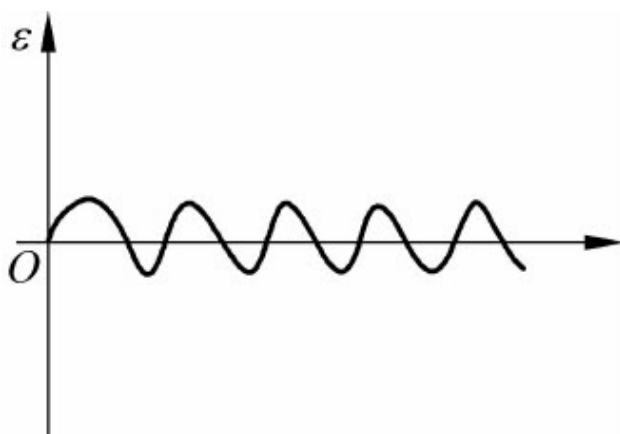


图6-4 闵氏时空零点能

这种能量以最简单的形态出现，那就是具有黑体谱的热辐射状态。因此，伦德勒观测者觉得自己浸泡在热浴之中（图6-5）。

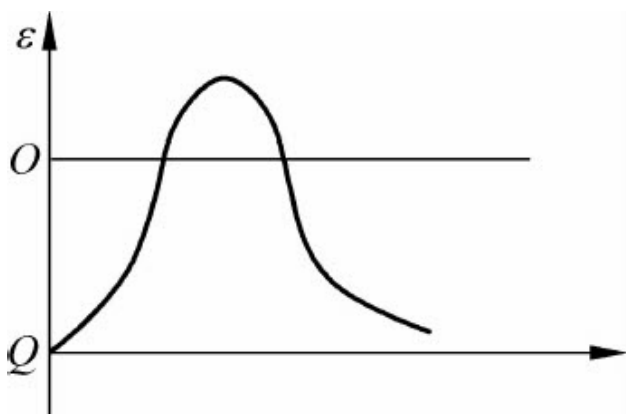


图6-5 伦德勒时空，真空能量的零点下降到Q点，闵氏真空的零点能以热能形式出现

霍金证明了黑洞有热辐射之后，安鲁很快认识到，自己发现的效应

与黑洞热辐射有相同的本质。因此许多人把霍金辐射与安鲁效应一起，称为霍金-安鲁效应。

黑洞的负比热

令人惊讶的是，黑洞的比热容是负的，因而在一般情况下，黑洞不可能与外界达到稳定的热平衡。这是因为史瓦西黑洞的表面引力为

$$\kappa = \frac{1}{4M} \quad (6.11)$$

这意味着黑洞的温度

$$T = \frac{1}{8\pi MK_B} \quad (6.12)$$

此式表明黑洞的温度与它的质量成反比。它放出热辐射后，质量 M 减小，温度 T 升高；它吸收外界辐射或物质后， M 增大，温度 T 反而降低。所以它的比热容是负的，以前我们几乎没有见过负比热容的物质。

通常的物质比热容都是正的，吸热后温度升高，放热后温度降低。所以通常物质系统一旦与外界达到热平衡后，平衡就会是稳定的。

如果平衡后由于热涨落外界温度升高一点，热量就会从高温的外界流入系统，正比热容导致系统温度升高，与外界重新达到热平衡。

如果热涨落使系统温度略高于外界，就会有热流从系统流出，系统温度就会下降，也重新与外界达到热平衡。我们通常见到的正比热容物质都是这样，系统与外界的热平衡因而是稳定的。

但是负比热容的黑洞与外界的热平衡不会是稳定平衡。假设一个黑洞与外界达到了热平衡，由于热涨落，外界温度可能会比黑洞升高一点或降低一点。

如果外界温度升高一点，就会有热流流入黑洞，黑洞质量会增加，由于比热容为负，黑洞温度反而降低，洞内外温差加大，高温的外界会有更多的热流进入黑洞，黑洞质量会进一步增大，温度会进一步降低。黑洞与外界的热平衡被打破，黑洞将不断吸收外界物质和热辐射，不断扩大降温，这一过程将永远持续下去。

如果黑洞与外界处于热平衡时，涨落使黑洞温度略高于外界，黑洞热辐射将增加，有热流从黑洞流出，黑洞的质量会减小，负比热容使它的温度进一步升高，与外界的温差扩大，黑洞热辐射将继续增加，输出的热流更大，黑洞又进一步升温，这将导致黑洞温度不断升高，与外界的热平衡彻底被破坏，最后小黑洞会爆炸而消失。

仔细研究表明，在某些特殊情况下，黑洞与外界可以形成稳定的热平衡。一种情况是把史瓦西黑洞放到一个绝热的盒子中，设黑洞质量为 M ，如果达到热平衡时，盒子中热辐射气体的质量不超过黑洞质量的四分之一时，即不超过 $\frac{1}{4}M$ 时，黑洞与外部气体之间的热平衡可以是稳定的。

另一种特殊情况是，研究发现带电史瓦西黑洞（即R-N黑洞）有一个相变点，位于

$$Q^2 = \frac{3}{4}M^2 \quad (6.13)$$

处。当

$$Q^2 < \frac{3}{4}M^2 \quad (6.14)$$

时黑洞的比热容为负，与外界的热平衡不稳定。当

$$Q^2 > \frac{3}{4}M^2 \quad (6.15)$$

时，黑洞的比热容会变成正的，这时黑洞就可以与外界处于稳定的热平衡状态了。

第七章 黑洞的信息佯谬

信息疑难

从黑洞概念产生伊始，就有人指出黑洞存在信息疑难。根据无毛定理，黑洞外部的观测者失去了组成黑洞的物质的几乎全部信息，只能探知黑洞的总质量、总电荷和总角动量，黑洞内部的其他信息完全探测不到了。既不知道黑洞是在什么时候由什么形态的物质形成的，也不知道进入黑洞的物质处在什么状态。对于黑洞外的人，这些信息都失去了，似乎信息不再守恒。

不过，这时问题还不大，只不过信息锁在了黑洞内部，藏在了视界里面，但它们依然存在，并没有在宇宙中消失，从整个宇宙的角度看，信息依然守恒。

霍金辐射发现以后，这个问题就严重了。由于比热容是负的，黑洞温度会在热辐射过程中不断升高，直至整个黑洞爆炸消失。由于黑洞热辐射是标准的黑体辐射，几乎带不出任何信息，所以在原来落入黑洞的物质化为霍金辐射（标准的黑体辐射）的同时，它们所带的信息也全都从宇宙中消失了，信息真的不守恒了。

重子数、轻子数是否守恒？

问题还不仅这一个，粒子物理中的重子数守恒、轻子数守恒等定律也被破坏了。形成黑洞的物质，可以是正物质，也可以是反物质。正物质由质子、中子、电子等组成，重子数（一个质子或一个中子的重子数为+1）和轻子数（一个电子的轻子数为+1）都是正的；反物质由反质子、反中子、正电子等反粒子组成，重子数和轻子数均为负。

正物质或等量的反物质形成的黑洞从外界看来是相同的，它们的总质量、总电荷、总角动量都相同，它们的重子数和轻子数当然不同，但无毛定理告诉我们，外界观测者看不见这些毛。

黑洞热辐射时，这些物质化为霍金辐射跑出来，而按照目前的认识，霍金辐射中正粒子和反粒子的产生概率是相同的，与形成黑洞的物质形态无关。这样，重子数守恒定律和轻子数守恒定律都被破坏了。所以，考虑了霍金辐射的黑洞理论，信息守恒被破坏，重子数守恒和轻子数守恒也被破坏了。

不过，重子数守恒和轻子数守恒是否是自然界的严格定律有人有怀疑。李政道先生指出，严格的守恒律往往对应补偿场，在严格的全局对称性被破坏，化为局域对称性的时候，会产生补偿场，即规范场。例如，电荷守恒对应电磁场，能量守恒和动量守恒对应引力场。

而重子数守恒和轻子数守恒相应的对称性，却没有对应的补偿场。因此，有人质疑重子数与轻子数是否严格守恒。所以黑洞理论破坏重子数守恒和轻子数守恒还不是最严重的问题。

么正性与概率守恒

然而，信息不守恒造成的后果是严重的。它使得量子理论中的概率不再守恒，么正性被破坏。么正性是现代场论的基石之一，所有的量子过程的演化都是么正的。如果么正性不成立，整个量子场论都需要做大的改动。而且，概率不守恒也是让人难以接受的。

所以，包含霍金辐射的黑洞理论，在理论物理学家中引起了轩然大波，在相对论专家和量子论专家之间展开了一场大辩论。

霍金打赌

1997年，相对论专家霍金和索恩（著名的时间机器专家），与量子物理专家普瑞斯基开玩笑打赌，霍金和索恩认为黑洞过程信息不守恒，普瑞斯基认为信息守恒，谁输了谁给对方订一年棒球杂志。

霍金、索恩：

黑洞中的信息丢失了 1997 年

普瑞斯基：

黑洞中的信息不会丢失，
会逸出或残留

霍金经常和别人开玩笑打赌，有一次和索恩为某个问题打赌，谁输了谁给对方订一年黄色杂志。后来霍金宣布自己输了，让助手买了一年的黄色杂志送到对方的家中，引得索恩夫人大为不满。

这次，霍金与索恩认为落入黑洞的物质最后化为了霍金辐射，信息肯定不守恒。普瑞斯基则认为信息守恒十分重要，信息应该守恒。以普瑞斯基为代表的许多量子论专家认为，霍金辐射未必是标准的黑体辐射，可能有一部分信息会被带出黑洞。另外，黑洞虽然辐射时温度不断增高，但很可能在某一高温下出现某种“量子效应”，使霍金辐射戛然而止，信息作为炉渣保留在残余的黑洞里面。

2004年，霍金突然宣布自己输了！他在爱尔兰举行的一次国际相对论大会上作了一个报告，但这只是一个高级科普报告，没有给出具体的数学物理推演。他宣布自己输的理由是，以前把黑洞想象得太理想了，真实的黑洞过程，应该信息守恒。

索恩则强调，这事不能由霍金一个人说了算，自己不承认输了。霍金没有买到棒球杂志，买了一年板球杂志，让助手交给了普瑞斯基。普

瑞斯基表示，没有听懂自己为什么赢了。

这件事当时轰动世界。一些记者对到会的相对论专家作了民意调查，大部分人认为信息应该守恒，但也有相当一部分专家认为信息不守恒。

霍金当时表示，他不久将发表严格的数学物理论文来阐述自己为何改变观点，但我们至今也没有看到他发表这方面的学术论文。

维尔塞克与帕瑞克的证明

霍金为什么会改变看法呢？这是因为，当时许多量子论专家作了证明黑洞过程信息守恒的论证，特别是诺贝尔奖获得者维尔塞克及其学生帕瑞克的工作，这些工作影响了霍金的思想。维尔塞克是研究强相互作用的，因为对量子色动力学的贡献获得过诺贝尔物理学奖。他认为信息应该守恒，指导帕瑞克对黑洞热辐射作了深入研究。

他们认为，以前有关霍金辐射的证明，都忽略了辐射对黑洞的反作用。例如，考虑黑洞辐射一个光子，以往的证明都没有考虑光子射出后黑洞半径会缩小，他们认为，这一“缩小”虽然微乎其微，但影响很大。正是这一“缩小”使辐射出的光子带出了一点点信息，恰是这一点点信息保证了量子演化的么正性，保证了信息守恒。他们发表了论文，对自己的观点作了严格证明。他们的意见是有道理的，以往所有对霍金辐射的证明，都没有考虑辐射导致的黑洞半径收缩。

我们读过霍金和其他相对论专家证明黑洞热辐射的不少论文，而且自己也做过很多黑洞热辐射的研究。不随时间变化的、旋转、带电的黑洞是最一般的稳态黑洞，这类黑洞的霍金辐射中包含电子、质子等自旋为 $1/2$ 粒子的证明，就是首先由我们这个组，在刘辽教授和许殿彦教授带领下给出的。作这些证明时，我们同样没有考虑霍金辐射与黑洞半径收缩的相互影响。

为什么大家的证明都没有考虑霍金辐射导致黑洞半径的收缩呢？因为这一收缩实在太小了。太阳质量的黑洞，半径才3千米，射出一个粒子，你说质量能有多大变化？黑洞半径能收缩多少？所以大家都忽略了这一收缩。

帕瑞克与维尔塞克的工作很漂亮，他们严格证明了自己的猜测：正是这一忽略导致了信息丢失。他们在自己的论文中考虑了这一影响，最后严格证明了霍金辐射过程保持么正性，热辐射带出了一点信息，从而信息依然守恒。

黑洞表面的势垒

他们还指出，霍金虽然在解释自己的黑洞热辐射理论时，强调这种辐射是一种“量子隧道效应”，是洞内粒子通过逆时运动，穿越黑洞视界处的势垒的结果。但是从霍金的证明文章中并看不出势垒的存在，更没有给出势垒的具体情况和位置。维尔塞克和帕瑞克声称他们的文章给出了势垒的具体情况和位置。

隧穿过程的图像

他们把粒子理想化，看作一个球面波（S波），粒子（S波）射出黑洞前，视界位置在 r_{in} 处，粒子射出后，黑洞收缩，视界收缩到 r_{out} 处

（图7-1）。这时粒子（S波）已处于黑洞外了。他们认为，势垒仅仅在粒子射出的过程中出现，位置就在 r_{in} 与 r_{out} 之间。实际上，粒子射出的过程，就好像粒子不动，势垒向内移动的过程，从 r_{in} 处移动到 r_{out} 处（图7-2）。

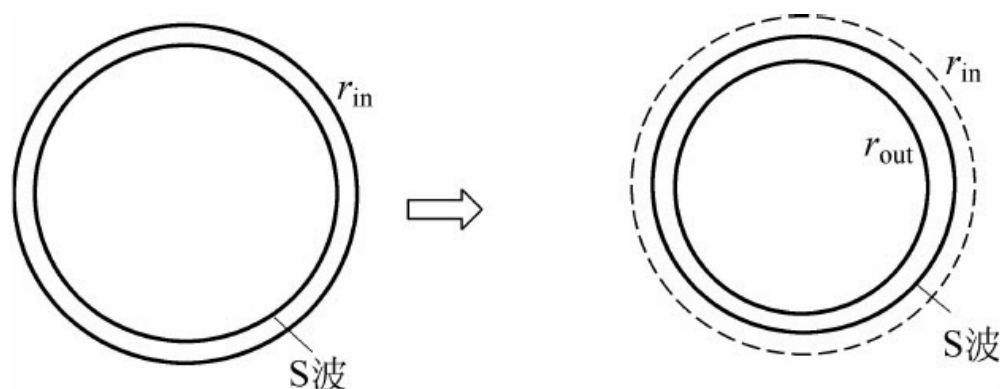


图7-1 隧穿过程：辐射使黑洞收缩

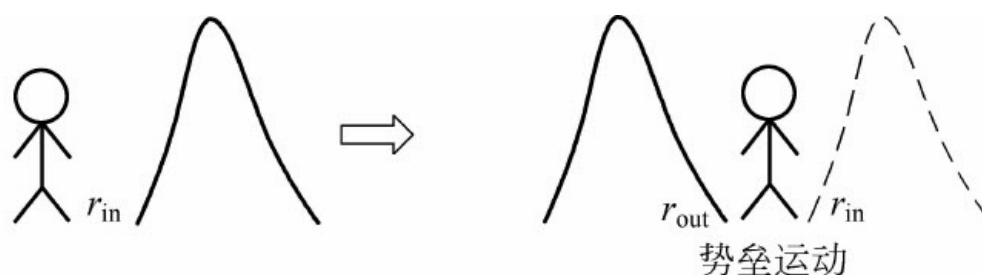


图7-2 隧穿示意图：粒子（图中小人）向外穿过势垒，等价于粒子不动，势垒向内移动

维尔塞克和帕瑞克等众多理论物理学家的工作，都似乎支持了黑洞过程信息守恒的观点，他们的努力促使霍金重新考虑自己的观点。他反思后认为，自己以前把黑洞想象得太理想化了，真实的黑洞过程比较复

杂，信息不会被绝对地锁在黑洞内部，也不会热辐射过程中真正丢失。于是他改变了自己的观点。霍金态度的转变，影响非常大，学术界似乎已经一边倒地认为黑洞过程不会破坏信息守恒。

然而，仍有一部分相对论研究者（例如索恩等人）不相信黑洞过程会信息守恒。

猜测：不存在信息守恒定律

我和我的学生张靖仪讨论了这个问题。我们认为，物理学中有能量守恒、动量守恒、电荷守恒等许多守恒定律，并没有公认的信息守恒定律。

信息论专家们认为信息可以视为“负熵”，包括霍金在内的许多物理学家也承认这一点。大家知道，“熵”恰恰是个不守恒的量，热力学第二定律的精髓就是认为熵不守恒。

孤立系统和绝热系统（与外界不存在热交换的系统），如果不处在平衡态，其中的熵只会增加，绝不可能减少。一般说来，自然界发生的过程（除去理想化的可逆过程之外）都是熵增加的过程。

所以，如果信息真的是“负熵”，在自然过程中，它肯定会不断减少，绝对不会守恒。因此，我们认为，没有理由认为“黑洞过程”必定信息守恒。而且，我们认为黑洞过程应该有信息丢失。

黑洞过程信息不守恒

前面我们已经谈到，黑洞的比热容是负的，这使得黑洞不可能与外界形成稳定的热平衡。即使黑洞在某一瞬间与外界达到了热平衡，温度相等，热涨落一定会破坏这一平衡，使黑洞和外界形成温差，而且这种温差会不断扩大。作为霍金辐射的热流出射过程，以及落入黑洞的辐射构成的热流入射过程，都将由于温差的存在而成为“不可逆过程”。热力学第二定律告诉我们，“不可逆过程”是熵产生的过程。如果信息是负熵，不可逆过程就是信息丢失的过程。所以，黑洞的辐射和吸收过程，都将是信息丢失的过程，信息肯定不会守恒。

帕瑞克证明的漏洞

那么，维尔塞克和帕瑞克怎么会证明了黑洞的霍金辐射信息守恒呢？我们猜测他们在证明过程中可能假设了霍金辐射是一个“可逆过程”。“可逆过程”中没有熵产生，熵是守恒的，信息自然也会是守恒的。

我们对帕瑞克等人的工作进行了详细研究。他们是以史瓦西黑洞为例作的证明，证明了这种球对称的静态黑洞在辐射静质量为零的粒子时信息守恒。为了弄清楚他们证明过程中是否假定了霍金辐射是一个可逆过程，我们把他们的工作推广到了更一般的情况。

我们先把帕瑞克等人的工作推广到了一般的稳态黑洞，包括黑洞转动和带电的若干情况；又把黑洞辐射的粒子从静质量为零的情况推广到静质量不为零的情况，以及辐射粒子带电的情况，等等。我们发现，对于这些更一般的情况，帕瑞克等人的证明，都有效，都成立，么正性都能保证，信息都守恒。

然而，我们也注意到一个情况：他们的所有证明都需要使用一个热量表达式

$$\Delta Q = T ds \quad (7.1)$$

式中 ΔQ 为系统与外界交换的热量， T 为温度， ds 为过程中熵的变化。

热力学告诉我们，此公式仅在可逆过程中成立，对于不可逆过程则不成立。一般说来，对于不可逆过程，应有

$$\Delta Q < T ds \quad (7.2)$$

因此，我们认识到，帕瑞克与维尔塞克等人的这一工作，在数学上是严格的，而且是普适的，对于各种黑洞、辐射各种粒子都成立。但他们都用了黑洞的霍金辐射过程是可逆过程的假定。

因为黑洞的比热容为负，不可能与外界达成稳定的热平衡，黑洞与外界必定存在温差，不管是作为黑洞出射热流的霍金辐射，还是黑洞吞食外部热流的吸积过程，都是不可逆过程，熵一定会增加，因此也就不可能信息守恒。

所以，我们的研究表明，维尔塞克等人的工作还不能认为已经证明了黑洞辐射过程信息守恒。有关黑洞过程是否信息守恒的争论还将继续下去。

我和我的学生们对信息守恒问题的研究工作，已在国内外很多杂志上发表。其中一篇（张靖仪，赵峥）还于2008年获得了汤姆孙-路透“科学前沿——中国卓越研究奖”。

黑洞中的信息当真不会逸出吗？

不过，霍金认为以往“把黑洞过程过于理想化”了的观点也是有道理的。真实的黑洞过程是否会有部分信息逸出呢？虽然总体来说黑洞过程信息不会守恒，但落入黑洞的物质的全部信息就真的一点都跑不出来吗？

我们认为，也许会有部分信息从黑洞中逸出。现在我们给出两种信息从黑洞中逸出的可能途径。

动态黑洞的启示

以往对于黑洞辐射的研究，都局限于稳态黑洞，即不随时间变化的黑洞。这是因为人们认为，辐射出的粒子的能量远小于黑洞能量，辐射过程对黑洞的影响完全可以忽略。另一方面，如果考虑黑洞的变化，研究中的困难会大大增加。

不过，我们前面介绍帕瑞克等人的工作时，就看到忽略辐射过程对黑洞的反作用，是值得推敲的。他们的工作虽然未能证明黑洞辐射过程信息守恒，但他们向人们明确显示了不应轻易忽略辐射过程对黑洞的影响。

而且，真实的黑洞都是在不断变化的，不断会有物质落入黑洞。特别是黑洞周围往往形成吸积盘，盘中的物质会大量落入黑洞，引起黑洞半径的变化。另外，帕瑞克只考虑了一个出射粒子自身对黑洞的影响，真实的黑洞辐射绝不会是一个粒子、一个粒子地往外跑，往往是大量粒子一起往外涌。特别是高温下的黑洞，出射能流（热辐射流）会引起黑洞半径的明显变化。

真实的黑洞都处在不断吸积和辐射的过程中，因而质量和半径都处在不断变化中，所以真实的黑洞都是动态的，不是稳态的。黑洞的这种非稳态状况，会不会对信息逸出产生影响呢？会不会有信息从动态黑洞的辐射过程中跑出去呢？

动态黑洞热性质的探索

对动态黑洞的研究比对稳态黑洞的研究困难得多。20世纪80年代后期开始出现一些这方面的研究，但都不够成熟，主要是动态黑洞的半径（视界位置）和温度都很难确定，不存在计算动态黑洞视界位置和温度的有效方法。人们用了不少近似，收效不大，而且研究对象仅局限于球对称黑洞。

1990年前后，我和我的学生戴宪新、罗志强、朱建阳、黎忠恒等人对动态黑洞的霍金辐射进行了深入研究，创建了一个研究动态黑洞热辐射的成功方法，不用事先假定黑洞的半径、温度或表面引力，就可以证明黑洞产生黑体辐射，并反过来自动定出黑洞的半径（视界位置）与温度。而且，我们可以逐点计算黑洞表面的温度，这是此前任何人都做不到的。

我们的方法是普适的，不仅可用于球对称的动态黑洞，而且可以用于非球对称的动态黑洞。我们证明，对于非球对称的动态黑洞，其表面温度不是一个常数，不仅随时间变化，而且随方位变化，也就是说这类黑洞表面各点温度不同，存在热流。我们的工作已在国内外许多杂志上发表，但没有引起足够的注意。

近年来，国外的一些相对论专家用另外的方法研究了动态黑洞的热性质，但局限于球对称的动态黑洞。

动态黑洞信息逸出的可能性

我们现在想谈一下对动态黑洞过程信息是否守恒的思考。

对于球对称的静态黑洞，它的三个特征曲面，即事件视界、表观视界和无限红移面是重合的。对于非球对称的稳态黑洞，其事件视界与无限红移面分开，中间出现能层，但表观视界仍然与事件视界重合。

对于球对称的动态黑洞，其表观视界 r_{AH} 与无限红移面 r_{TLS} 重合，但事件视界 r_{EH} 与它们分开了（图7-3），这时在事件视界与表观视界之间出现一种新的能层——量子能层。

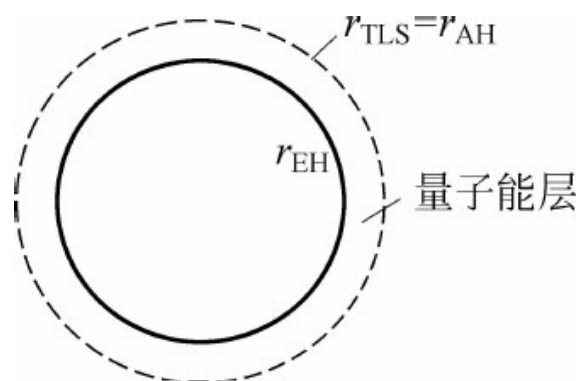


图7-3 动态黑洞的视界和量子能层

黑洞的霍金辐射产生自表观视界还是事件视界，国外的一些文献认为产生自表观视界。但我们的研究表明，霍金辐射产生于事件视界，而不是表观视界。与国外的主流方法不同，我们没有事先假定辐射产生自哪一个视界，而是在用我们自己的方法导出霍金辐射的过程中，自动确定出辐射产生自事件视界，而不是表观视界，并自动定出热辐射的温度。

我们证明了产生自事件视界的霍金辐射具有严格的黑体谱。问题是当此辐射穿越量子能层时是否有变化？是否偏离黑体谱，从而带出信息？

如图7-3所示，表观视界位于事件视界之外，量子能层中时间箭头的方向仍然指向黑洞内部，因此粒子可能无法经典地穿过量子能层，可

能还要再通过量子效应（如隧道效应）才能射出。量子能层处很可能存在势垒，此势垒是否使辐射偏离黑体谱，带出信息，值得探讨。

此外，我们在证明事件视界产生霍金辐射时，使用了粒子的动力学方程（如克莱因-高登方程，狄拉克方程等）在视界处的渐进方程。渐进方程的使用是否引入了某些近似，丢失了某些东西，也是值得进一步推敲的。

因此，对于动态黑洞的霍金辐射会否带出部分信息，还是值得进一步研究的，目前下定论尚为时过早。

时空扰动引来的启示

在用整体微分几何研究时空的时候，有人注意到外界对时空的扰动，有可能引起光锥的变化，如图7-4所示。在未发生扰动时光锥是 45° 角张开。光锥内部是A点的因果过去（下半光锥）和因果未来（上半光锥）。处于A点的粒子若静质量不为零，则只能到达上半光锥的内部，若静质量为零，则只能沿上半光锥的光锥面运动。无论该粒子是否静质量为零，都不可能跑到光锥外面去，因此不可能到达B点。

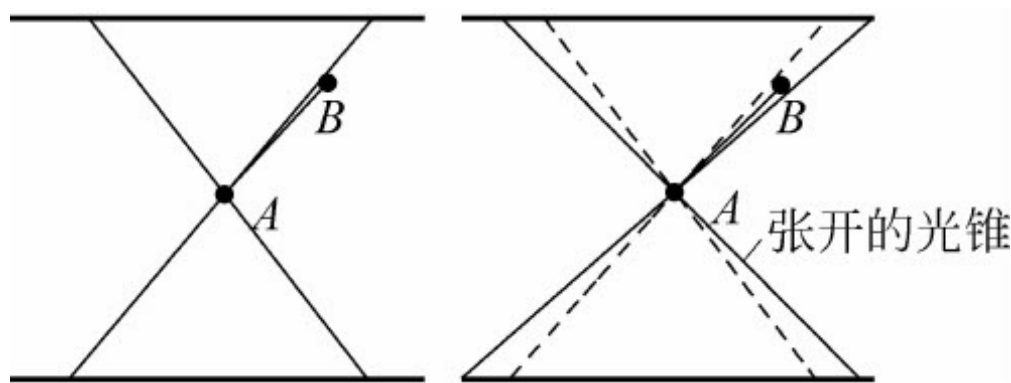


图7-4 时空扰动导致的光锥变化

但是，如果时空受到扰动，光锥有可能变胖，如图7-4右边所示，这时B点有可能被包在光锥的内部，处于A点的“因果未来”，因此位于A点的粒子就可以跑到B点。

如果此后光锥又在扰动下变瘦，已跑到B点的粒子又位于光锥之外了。这样，在时空扰动下，A点的粒子跑到了光锥以外的B点，那本来只有超光速粒子才可以到达的地方。现在，粒子并没有超光速，却已经到达了那里。

黑洞的表面（事件视界）就可以视为光锥面，位于黑洞内部或黑洞

表面（例如A点）的粒子本不可能跑到黑洞外部的B点。但由于时空受到了扰动，光锥时而张大，时而缩小，则位于A处的粒子就有可能跑到黑洞外部去，并带出黑洞内部的一些信息（图7-5）。

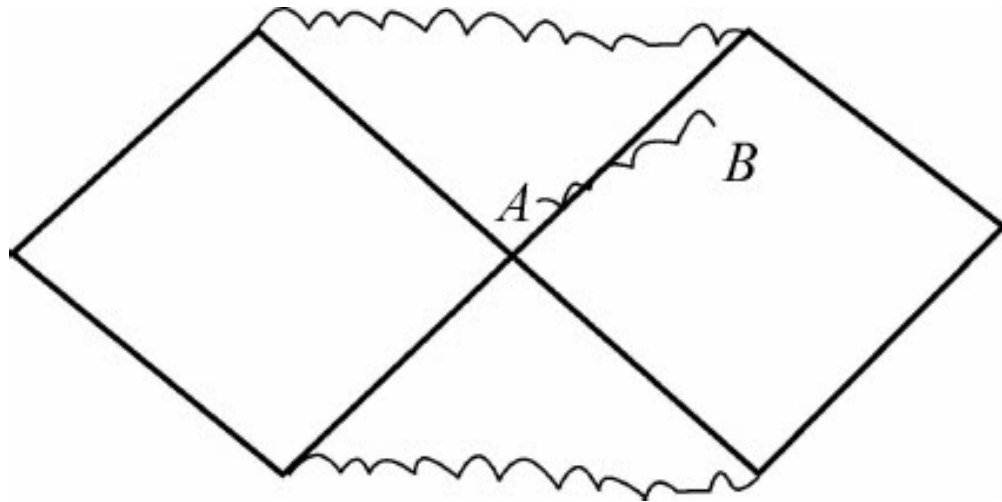


图7-5 黑洞视界附近的时空扰动造成洞内粒子和信息的逸出

特别是近极端黑洞的情况，单向膜区已薄到几乎只剩一层膜，上述时空扰动非常可能使洞内粒子及其所带信息逸出到洞外。

结论与讨论

我们可以得到几点结论：

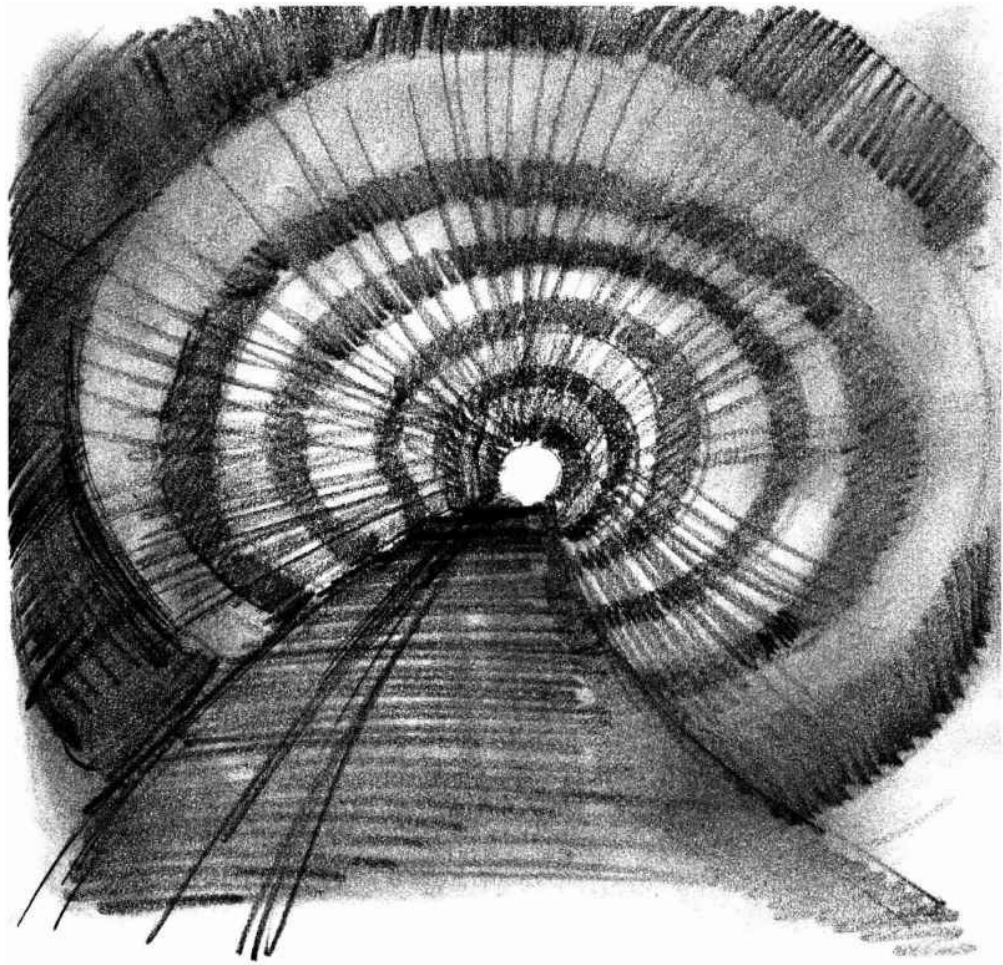
第一，物理上不存在信息守恒定律，如果信息是负熵，则信息应该不守恒。不应企望黑洞过程（如吸收或辐射等过程）信息守恒。

第二，维尔塞克与帕瑞克关于霍金辐射信息守恒的证明，在数学上是正确而严格的，所用的方法对各种黑洞和各种粒子均适用。

但是，他们的证明过程中假定了过程可逆，然而真实的黑洞过程肯定是不可逆过程。在不可逆过程中得不到他们希望得到的结论。所以，他们的工作未能真正证明黑洞过程信息会守恒。

第三，霍金认为以前把黑洞过程看得太理想化了，真实的黑洞过程中，可能会有部分信息逸出，也可能会有部分信息永久锁死在黑洞内部。这些想法可能是对的，但这并不意味着信息一定守恒。

第四，我们谈到了黑洞中信息可能逸出的另外两点考虑：动态黑洞热辐射有可能偏离黑体谱；黑洞视界附近时空涨落，有可能导致一些信息逸出。然而，有关问题还需要进一步研究。



绘画: 张京

第八章 奇点——时间有无开始与终结

奇点定理的哲学意义

现在我们把注意力转向黑洞内禀奇点的研究。这种奇点时空曲率发散，而且此类发散不能通过坐标变换来消除，因而它表示时空本身存在奇异，也就是说存在“病态”。令人意外的是，对内禀奇点的研究把科学家们引向了“时间是否有开始和终结”的探讨。

时间有没有开始和结束？千百年来，许多伟大的思想家对此进行过深入的探索，但那都是些哲学家和神学家，而且有关的探讨都局限在哲学分析、神学研究和主观猜测上。从20世纪60年代开始，物理学介入了这一问题的研究。其标志是彭罗斯和霍金提出的奇点定理（或称奇性定理），该定理概括并超出了关于宇宙开端和终结的研究。

奇点定理可粗略表述为：只要广义相对论成立，因果性良好，有物质存在，就至少有一个物理过程，其时间存在开始或存在结束，或既有开始又有结束。这一数学定理在物理学和哲学上的重大意义是不言而喻的。遗憾的是，到目前为止，它还没有引起哲学界的注意，科学界对它的重视也远远不够。

下面，我们将对奇点定理及其造成的困难作简要的介绍，并讨论其可能引发的重大科学与哲学进展。

内禀奇点与坐标奇点

广义相对论诞生不久，人们就发现爱因斯坦方程的解（即满足广义相对论的时空）普遍存在奇异性（奇点或奇环等）。奇异性有两类，一类是内禀奇异性，是时空本身存在问题，表现为时空曲率发散，而且这种发散不能通过坐标变换加以消除。例如，

球对称黑洞（史瓦西黑洞）的“中心”奇点，转动黑洞内部的奇环，大爆炸宇宙的初始奇点（图8-1中的O点），大塌缩宇宙的大挤压终结奇点（图8-1中的p点）等，都属于这类奇异性。

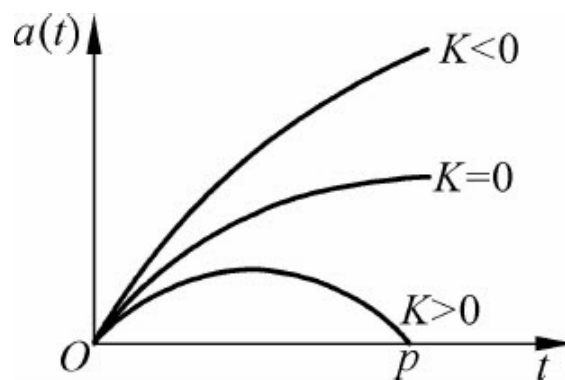


图8-1 宇宙的初始奇点和终结奇点

另一类是坐标奇异性，这种奇异性不是时空本身存在问题，而是由于坐标系选择不当引起的，可以用坐标变换加以消除。只存在坐标奇异性的地方，时空曲率正常，并不出现发散。当然具有坐标奇异性的地方，往往也有物理意义，例如前面谈到的各种黑洞的表面（事件视界），都存在坐标奇异性。

不过，下面我们探讨的都是内禀奇异性。为了讨论方便，以后我们把出现内禀奇异性的地方（奇点、奇环等），统称为奇点。

栗弗席兹与卡拉特尼科夫的失误

20世纪70年代对奇点问题的深入研究，起源于苏联物理学家的的工作。当时朗道已经去世，他的助手栗弗席兹与另一位物理学家卡拉特尼科夫联手，对广义相对论中的奇点问题展开了研究。

把他们吸引到这一问题上来的原因是，他们注意到当时已知的广义相对论场方程的解，除去作为真空的闵可夫斯基时空和德西特时空（常曲率时空）之外，一般都存在内禀奇点，奇点处时空曲率发散，因而是物理理论无法了解的地方，它随时可能产生无法预测的信息。环形奇点（如转动黑洞中的奇环）的附近，还会出现“闭合类时线”，沿这类曲线生活运动的人，会回到自己的过去。这简直令人不可思议。因此他们觉得，真实的物理时空不应该存在奇点。

那么，为什么满足广义相对论的时空普遍存在奇点呢？他们认为，这是因为人们在求解爱因斯坦场方程时，把时空的对称性假设得过于理想所致。

他们想，当一个星体做标准的球对称塌缩时，所有的星体物质都球对称地下落，因此最终都挤到球心，形成奇点，例如史瓦西黑洞中心的奇点。

但是真实的星体塌缩不可能是标准而严格的球对称塌缩，所有各个方向塌向球心的物质会在中心附近交叉而过，最终不会形成奇点。

克尔黑洞中的奇环，也是人们把星体塌缩想得太理想所致，当旋转的星体做标准的轴对称塌缩时，才会形成奇环。但真实的旋转星体，塌缩时不会严格轴对称，塌缩的星体物质也会相互碰撞，最终“擦肩而

过”，不会形成奇环。

总之，他们认为奇点和奇环的出现，都是因为人们把星体塌缩时的对称性想得太好所致。真实的塌缩过程不可能保持如此严格的对称性，所以奇点和奇环都不可能真正形成。

人们为什么要把物质和时空的对称性设想得那么完美呢？这是因为广义相对论的场方程过于复杂，求解起来非常困难。把对称性设想得越好，场方程越可能化简，越容易求解。所以，场方程的复杂性迫使科学家们把时空的对称性想象得尽可能好，这样，他们才求出了场方程的一些解。卡拉特尼科夫与栗弗席兹认为，令人遗憾的是这也导致了奇点的出现。

他们进行了较为深入的研究，最后得到一个结论：奇点并非广义相对论的必然结果，奇点的出现是把时空和物质的对称性设想得太好所致。真实的时空和物质分布不可能保持严格的对称性，所以真实的时空中不应该存在奇点。

后来，栗弗席兹与卡拉特尼科夫的这一结论被证明有错误，他们本人也认识到了。

彭罗斯的创新思维

彭罗斯不相信他们的结论，认为他们的证明过程有误。

彭罗斯指出，奇点是广义相对论理论固有的“毛病”，满足广义相对论的真实时空中都不可避免地存在奇点。

经过仔细思考与研究，彭罗斯和霍金证明了奇点定理。这个定理可粗略表述为：只要爱因斯坦的广义相对论正确，并且因果性成立，那么任何有物质的时空，都至少存在一个奇点。

更加值得注意的是，彭罗斯和霍金在提出并证明奇点定理的过程中，对奇点概念进行了重新认识，提出了极其重要的新思想：奇点应该看作时间的开始或终结！

这就是说，他们的奇点定理证明了一定存在时间有开始和终结的过程，这使得物理学介入了“时间有无开始和终结”的探讨。

彭罗斯与霍金等人对于奇点的这一认识，来源于对宇宙和黑洞的研究。在大爆炸宇宙模型中，宇宙与时间一起诞生于时空曲率发散的初始奇点；对于其中的大塌缩结局，宇宙与时间又一起终结于时空曲率发散的大挤压奇点。

另一方面，广义相对论告诉我们，黑洞内部的时空坐标要发生互换，原来的时间 t 成为空间坐标，而径向坐标 r 则成为时间坐标。所以黑洞内部的等 r 面不再是球面，而成为等时面。

对于黑洞，时间方向指向 $r=0$ 的奇点处。这样，等 r 面成为“单向膜”，任何进入黑洞的物质只能向 r 减小的方向运动，不能停留，也不可

能反向运动，而且没有任何力和任何物质结构能够抗拒这种运动。

这是因为，这不是一般的运动，而是一个时间发展的过程，什么力量都不能抵挡，不能不顺着时间方向前进。也就是说，任何物质都必须“与时俱进”。

黑洞内部整个是单向膜区，黑洞的边界（视界）是单向膜区的起点。进入黑洞的飞船和任何其他物质都将在有限的时间内穿越单向膜区到达奇点。值得注意的是，由于时空坐标互换， $r=0$ 现在不是黑洞的“球心”，而是时间的终点。这就是说，进入黑洞的飞船和宇航员的时间将在有限的经历中结束。

按照广义相对论，还可能存在白洞。白洞是黑洞的时间反演。它的内部也是单向膜区，只不过时间方向从奇点 $r=0$ 处指向视界处，所以它的单向膜的单向性与黑洞相反。需要强调的是，白洞内部的 $r=0$ 处，不是时间的终点，而是时间的起点。

奇点：时空中无法修补的洞

有奇点的时空，称为奇异时空。然而，如果有人把奇点从时空中挖掉，剩下的时空还能叫做奇异时空吗？彭罗斯和霍金认为即使把奇点挖掉，时空的根本性质也不会有变化，仍然是奇异时空。

然而，挖掉奇点之后，时空中就不存在曲率为无穷大的点了，因此，仅仅用“曲率无穷大”来定义奇点是有缺陷的。他们注意到，虽然人们可以把奇点从时空中挖掉，但挖掉之后总会留下空洞，那么时空中任何一条经过空洞的曲线（世界线）都会在那里断掉。

于是，彭罗斯和霍金建议，干脆把奇点从时空中“去掉”，认为它们不属于时空（图8-2）。粗略地说，干脆把它们看作时空中的“空洞”。但是任何一个正常点也都可以从时空中挖掉，形成空洞，时空中的曲线到达这样的空洞当然也会断掉，不过，这种空洞可以用数学方法补上，而奇点处的空洞则由于曲率发散不可能补上。

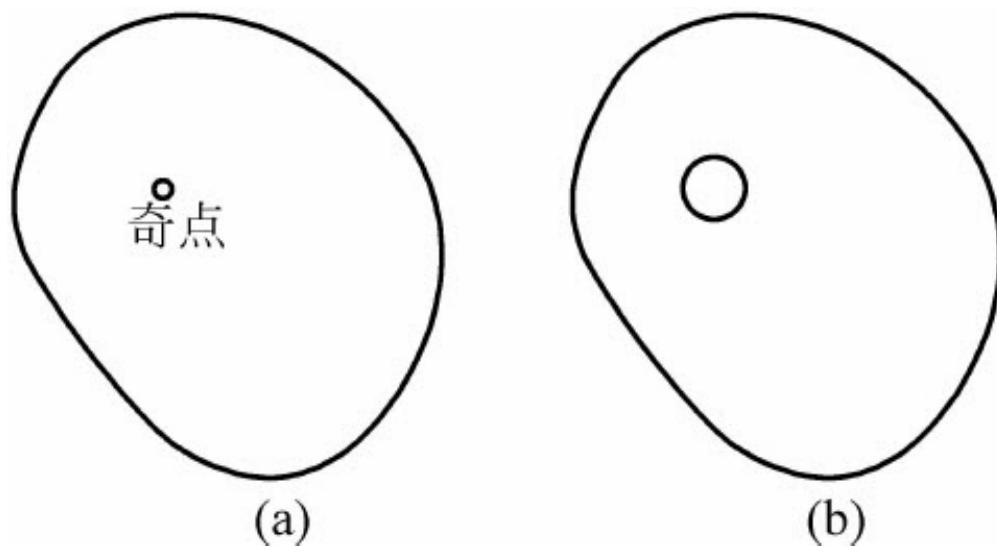


图8-2 (a) 奇异的时空及 (b) 挖掉奇点的奇异时空

奇点：时间的开始与终结

于是，彭罗斯和霍金这样去证明他们的奇点定理。证明时空中至少存在一条具有如下性质的类光（光速）或类时（亚光速）世界线：它在有限的长度内会断掉，而且断掉的地方不能用任何手段修补，以使这条世界线可以延伸过去。

类空（超光速）世界线不在他们的考虑范围之内，因为这样的曲线描述超光速运动，而自然界不存在超光速运动的粒子。

类光世界线描述光子运动，类时世界线描述低于光速的质点的运动，例如电子运动、火箭运动以及我们人类可以进行的任何活动。总之，光速或亚光速世界线描述自然界存在的一切实际过程。

相对论研究表明，时空中的亚光速世界线的长度，恰恰是沿此线运动的质点（或火箭、或任何物体和人）所经历的真实时间（称为固有时间）。所以，按照彭罗斯和霍金的观点，奇点就是时间过程断掉的地方。

奇点定理的实质内容是：在因果性成立、广义相对论正确而且有物质存在的时空中，至少有一个可实现的物理过程，它在有限的时间之前开始，或在有限的时间之后终结。也就是说，至少有一个物理过程，它的时间有开始，或有终结，或者既有开始又有终结。换句话说，至少有一个时间过程，它的一头或两头是有限的。

总之，奇点定理告诉我们，时间不都是无穷无尽的。黑洞的内部，有一个时间的“终点”，即黑洞的奇点。白洞的内部，有一个时间的“起点”，即白洞的奇点。膨胀宇宙的时间有一个起点（大爆炸奇点），脉

动宇宙的时间，则不仅有一个起点（大爆炸奇点），还有一个终点（大挤压奇点）。

奇点定理的前提条件是无可非议的。奇点定理的证明过程，依据了现代微分几何和广义相对论的研究成果，经过了不少专家的反复推敲。看来，奇点困难无法摆脱。奇点定理不仅确认了奇点不可避免，而且指出奇点困难反映了时间的有始有终性。

下面，我们先介绍一些基本概念，然后介绍证明奇点定理的思路。

测地线与仿射参量

由于一般的世界线反映的不一定是质点（或光子）的自由运动，可能有外力作用，即反映的不一定是纯“时空”性质，而且一般世界线不易找到合适的参量来表征“长度”，彭罗斯和霍金在研究奇点时把注意力集中到测地线（即“短程线”）这类特殊的世界线上。

测地线是直线在弯曲时空中的推广，它是不受外力（万有引力不算外力）的自由质点和自由光子在弯曲时空中的运动轨迹，反映的是时空本性，而且测地线有一种很好的参量可以反映长度，那就是仿射参量。

类时测地线（自由质点的轨迹）的仿射参量可以看做固有时间（即沿此测地线运动的观测者亲身经历的时间）。类光测地线（自由光子的轨迹）的仿射参量虽然不能看做固有时间，但仍能很好地描述光线的长度。

如果有一根非类空测地线（即类时或类光的测地线），在未来或过去方向上，在有限的仿射长度内断掉，不能再继续延伸，那么，这根测地线就被认为碰到了时空的“洞”。如果这个“洞”补不上（例如，曲率发散处的“洞”就补不上），那么它就是奇点。

严格说来，“洞”不一定是一个点，可能是一个区域，而且此区域不属于时空，甚至可能不属于流形，个别情况还不属于拓扑空间。

注意，现代广义相对论认为，奇点本身不属于时空。奇点与无穷远点均不属于时空，区别在于伸展到无穷远点的光线或类时线长度无限，即光线的仿射参量和类时线的固有时间趋于无穷；而伸展到奇点的光线和类时线长度有限，即仿射参量或固有时间有限。

时空的因果结构

分别满足下述5个条件的时空，具有不同的因果结构，因果性一个比一个好。

（1）编时条件：不存在闭合类时线。即一个人或一个质点不能随着时间前进，又转回自己的过去。

人或质点不管在三维空间中静止还是运动，在四维时空中都会描出一条亚光速的世界线——类时线。此类时线描述了人或质点的“历史轨迹”和“生活经历”。

如果此类时线闭合，就表明这个人或质点会回到自己的过去，影响过去的自己，从而改变自己过去的历史，造成因果性的破坏。编时条件就是禁止出现这种荒谬的情况。

（2）因果条件：不存在闭合因果线。即不仅没有闭合类时线，也没有闭合类光线。闭合类光线表示一条光线随着时间前进，会转回自己的过去，这同样会影响过去，破坏因果性。

（3）强因果条件：不存在闭合因果线，也不存在无限逼近闭合的因果线。

沿“无限逼近闭合的因果线”发展的人、质点或光子，不可能完全不影响自己的过去。

（4）稳定因果条件：在微扰下也不出现闭合因果线。即不存在闭合因果线，而且在对时空进行微扰的情况下，也不会导致原来不闭合的因果线闭合起来。

物质的存在和运动变化，不可能不对时空产生反作用，造成微扰。如果一有微扰因果线就闭合，这种时空的因果结构还是有问题的。

（5）整体双曲：时空存在柯西面。所谓柯西面是这样一张超曲面，时空中的任何一条因果线都必须与它相交，而且只交一次。

整体双曲时空被一张张柯西面充满。粗略地说，每张柯西面都可以看作该时空的一张“同时面”。而且，我们只要知道了任何一张柯西面上的全部数据，就可以知道整个时空的全部“历史”，所有的这些数据如何从过去演变而来，又如何继续发展演变。非整体双曲的时空都做不到这点。

所以许多人认为，真实的物理时空都应该是整体双曲的。整体双曲有可能是物理学对时空性质的要求。当然，对于这一点，学术界还有不同意见。

感兴趣的读者可参看梁灿彬与周彬合著的《微分几何与广义相对论》。

整体双曲的时空是因果性最好的时空。整体双曲的时空一定稳定因果，稳定因果的时空一定强因果，强因果时空一定满足因果条件，因果条件一定能推出编时条件。

闵可夫斯基时空和史瓦西时空都是整体双曲的。Reissner-Nordstrom 时空（即带电史瓦西时空）是稳定因果的。转动轴对称的克尔时空和克尔-纽曼时空（带电克尔时空）则因果性很差，连编时条件都不满足，在奇环附近存在闭合类时线，沿此类时线生存的观测者，将不断地返回自己的过去。

能量条件

真实的时空，不应该一点能量都没有。能量条件就是给出时空中能量和压强存在的情况。

(1) 弱能量条件

此条件表示能量密度 ρ 一定非负

$$\rho \geq 0 \quad \text{且} \quad \rho + p_i \geq 0 \quad (i = 1, 2, 3) \quad (8.1)$$

式中， p_i 为压强（应力）。

(2) 强能量条件

$$\rho + \sum_{i=1}^3 p_i \geq 0 \quad \text{且} \quad \rho + p_i \geq 0 \quad (i = 1, 2, 3) \quad (8.2)$$

式中， ρ 为能量密度， p_i 为压强（应力）。强能量条件是说应力不能太负。事实上，在绝大多数情况，应力都是正的，所以一般情况下强能量条件反而比弱能量条件弱。但是，存在应力为负的情况，这时，强能量条件就比弱能量条件强了。

(3) 主能量条件

$$\rho \geq |p_i| \quad (i = 1, 2, 3) \quad (8.3)$$

主能量条件是要求能量流不能超光速，且弱能量条件必须成立。从主能量条件可以推出弱能量条件。

共轭点

讨论奇点定理，需要用到测地线的共轭点的概念。我们来介绍一下共轭点。

测地线是直线在弯曲时空中的推广。大家都知道，平直时空中的两条直线只能有一个交点。弯曲时空则不同，两条测地线可能不只一个交点。例如地球表面的经线，它们都是大圆周，都是球面上的测地线（短程线），但它们在北极和南极各有一个交点，一共两个交点，这两个交点称为“共轭点”。

在一般的弯曲时空中，我们可以定义测地线汇，就是在时空的每一点画出一条测地线，它们汇成一束，在这束“线汇”中，每一点都有一根且只有一根测地线。

测地线汇中无限邻近的测地线，如果有两个交点，则称此二交点为共轭点（图8-3）。测地线汇，按照定义，是指过每一时空点有一根且只有一根测地线的情况。在测地线的交点，当然有两根以上的测地线。从这个意义上讲，共轭点是线汇的奇点，但它还不是时空的奇点。

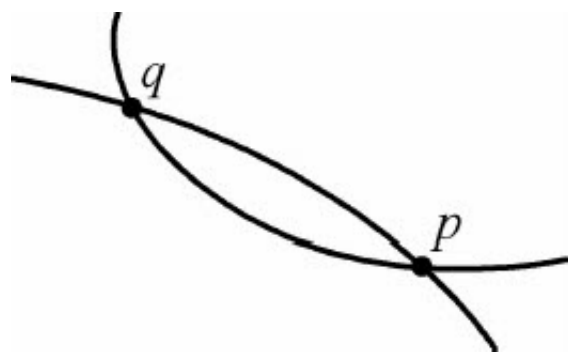


图8-3 线汇的共轭点

最长线：无共轭点的测地线

可以证明：不管是类时线还是类光线，两点之间长度取最大值的世界线一定是中间无共轭点的测地线。

前面我们曾经谈到，两点之间距离取极小值（或极大值）的线，一定是测地线（或称短程线）。这里我们又进一步看到，两点之间距离取极小（或极大）值的线，不仅必须是测地线，而且还要其上不存在共轭点，这是为什么呢？

我们用球面来举一个例子，大家就易于理解了。在地球表面，从北京到北极点最近的连线，一定是过这两点的大圆周（测地线），也即过北京的经线。

但是，从北京到北极点，沿这条经线有两条路可走，一条是从北京径直往北，到达北极点。另一条是从北京往南走，沿这条经线到达南极点，绕过南极后再继续向前，往北走，最终到达北极点。这两条路线都是沿同一条经线，也就是说沿同一个大圆周，同一条测地线。

显然，只有从北京出发径直往北的路径最短，从北京往南越过南极点再往北到北极的路径，不是最短的线。它们不都是大圆周吗？不都是测地线（短程线）吗？为何只有从北京径直向北的一条是最短线，往南的一条不是呢？

这是因为往南的一条上有共轭点，例如南极点就是与北极点共轭的点，共轭点的存在，使这段测地线不再是最近的距离。不经过南极点的那段经线上没有共轭点，所以只有从北京往北走，直达北极点的经线才是最短线。

奇点定理的证明思路

可以证明：

(1) 在强因果时空中，不一定有最长线；如果有，则一定是无共轭点的测地线。

(2) 在整体双曲时空中，一定有最长线，它一定是无共轭点的测地线。

另一方面，又可以证明：

如果广义相对论正确，强能量条件成立，并且时空中至少有一个存在物质的时空点，则测地线上在有限的仿射距离内必存在共轭点。

总之，因果性要求有最长线，即要求存在无共轭点的测地线。能量条件、广义相对论和物质的存在则要求测地线上一定有共轭点，而且是在有限的仿射距离内就出现共轭点。

如果时空同时满足上述因果性和能量条件，并存在物质，而且广义相对论正确，那就会导致矛盾的结论：测地线上既要有共轭点，又要无共轭点。

解决此矛盾的唯一出路是，测地线不能无限延伸，在出现共轭点之前，在有限的仿射距离内就断掉。也就是说，此测地线一定会遇到奇点，时空一定存在奇异性。

例如，有一个因果性很好的时空，在此时空中的任意两点 p 与 q 之间，必定有最长线，它一定是一条不存在共轭点的测地线，如图8-

4 (a) 所示。

如果此时空是广义相对论场方程的解，满足能量条件，且存在一点儿物质，那么从 p 到 q 之间的测地线一定存在共轭点，例如图8-4 (b) 所示，图中 r 点即是与 p 点共轭的点。

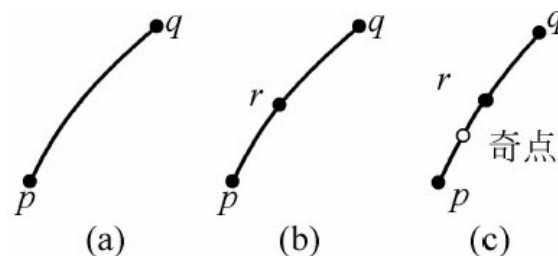


图8-4 共轭点与奇点

因果性要求 p 、 q 之间无共轭点，能量条件等又要求有共轭点。这个矛盾如何解决呢？唯一的办法是认为此测地线从 p 出发后，在抵达共轭点 r 之前碰到了洞，断掉了，即碰到了奇点（如图8-4 (c) 所示）。

由此，就证明了奇点定理：如果广义相对论正确，能量密度非负，时空不完全是真空，因果性良好，则时空一定存在奇点。

我们看到，史瓦西时空、克尔-纽曼时空、膨胀宇宙模型中都有奇点。不过，闵可夫斯基时空和德西特时空没有奇点，这是因为它们是完全的真空，没有任何物质存在。

通常确认时空奇点有两个步骤：一是证明有非类空测地线在某处不可延伸。二是证明反映时空曲率的标量在该处发散，这种发散使得该处的时空不能被修补，以使测地线延伸过去。

对第二个步骤应该做一点说明，在广义相对论中描述曲率的量，有的是张量，有的是标量。张量在坐标变换下会发生变化，它的某些分量如果在一个坐标系中发散，变到另一个坐标系就有可能不发散。但标量的值与坐标系的选择无关，如果在一个坐标系中发散，那么无论怎样变换坐标系，它都会保持发散。所以判断时空曲率是否出现发散，一定要用描述时空曲率的标量。

步入探讨奇点的殿堂

多数相对论专家相信，奇点困难是由于引力场没有量子化而造成的。奇点定理是经典广义相对论的结论。如果把引力场量子化，奇点困难可能会自动消失。

遗憾的是，引力场量子化的努力还远未成功。也有一些人由于种种原因不相信引力场量子化就能自然消除奇点困难。霍金本人则试图引入虚时间来化解奇点困难。而我在思考奇点之谜的过程中，走上了一条独特的道路。

1979年，中国科学界举行了纪念爱因斯坦诞生100周年的活动，在科学会堂的大会报告中，时任中国引力与相对论天体物理学会副理事长的郭汉英研究员介绍了彭罗斯与霍金的奇点定理。

这是我第一次注意到这个定理，它引起了我心里极大的震撼。物理学居然开始研究时间有没有开始和结束的问题，这一问题太重要了，它原本属于神学和哲学的范畴，现在居然进入了科学的范畴。

此后我开始注意有关奇点定理的各种介绍，翻阅了一些书籍和文献。后来梁灿彬先生在北师大用整体微分几何介绍现代广义相对论，把从芝加哥大学Wald教授那里学到的新东西介绍给中国相对论界。

我认真地听了梁先生的全部几个学期的讲座，记了8本笔记，并在与奇点定理有关的内容上反复琢磨。我经过不断阅读思考，逐渐加深了对这些问题的理解，产生了一些体会。

对奇点困难的另类思考

1980年前后我的主要注意力放在黑洞热性质的研究上，并不时在考虑热力学第零定律与“同时传递性”的关系。

当时，物理界都认识到时间的性质与热力学有关，但大家强调的只是“时间箭头”，只是热力学第二定律与时间流逝性的关系。

其实，物理界也早就知道能量守恒（热力学第一定律）与时间均匀性有关。只不过在谈论热力学与时间性质的关系时没有强调这一点，只是强调第二定律与时间箭头的关系。

在探索第零定律与时间测量之间的关系时，我注意到了热力学第二定律和第一定律也都与时间的性质有关。有时就想，热力学总共四个定律，前三个（第零、第一、第二定律）都与时间有关，第四个（即第三定律）是否也与时间性质有关呢？

我开始猜测，与第三定律有关的莫非就是奇点定理吗？“绝对零度的不可抵达性”真的和“时间有无开始与终结”有关吗？越想越觉得自己的推测可能有些道理，于是开始注意这方面的资料。

奇点伴随温度异常

我在对黑洞和奇点的长期研究中，注意到伴随奇点出现的一个重要物理特征：奇点总是伴随温度异常而出现。

在对黑洞和霍金热效应的研究中，我们注意到，这种效应依赖于坐标系（坐标温度）和观测者（固有温度）。凡是接触奇点的坐标系，都处于绝对零度或温度发散的状态；凡是有限温度的坐标系，都伸展不到奇点处。

例如，黑洞外部的观测者，接收不到来自黑洞内部的信息，更不可能看到位于黑洞内部的奇点或奇环。在他看来，任何落向黑洞的物体，或飞向黑洞的飞船，都会越走越慢，越来越发红（红移），越来越暗，最后冻结在黑洞的表面（事件视界）上，消失在那里的黑暗之中，永远也看不到它们落进黑洞。

因此，可以认为洞外观测者所用的坐标系（例如史瓦西坐标系）接触不到奇点。第六章的讨论还告诉我们，黑洞外面存在霍金辐射，因而外部观测者和他的坐标系处在有限温度的环境之中。已知的例子都告诉我们，有限温度的坐标系都接触不到奇点。

怎样才能使洞外的观测者能够看到奇点呢？办法是不断向黑洞注入电荷和角动量，使黑洞变成极端黑洞，内外视界重合， $r_+ = r_-$ ，单向膜区成为一张薄膜，这时再增加一点电荷和角动量，单向膜区就会消失，事件视界不再存在，奇点（或奇环）就会裸露出来，称为“裸奇异”，这时位于洞外的观测者就会看见奇点了。也就是说，洞外观测者所处的坐标系就接触到奇点了。但极端黑洞处在绝对零度，这就是说，一个坐标

系要接触到奇点，必须首先破缺热力学第三定律，使自己达到绝对零度，才有可能进一步接触到奇点。

有人用温度格林函数（松原函数）研究过黑洞的温度。研究发现，如果假定覆盖整个史瓦西时空的克鲁斯卡坐标系处在绝对零度，用零温格林函数来描述，则变到史瓦西坐标系后，此零温格林函数将自动变为有限温度的松原函数，其温度恰为史瓦西黑洞的霍金辐射温度。这样就证明了史瓦西时空处在有限温度，即霍金辐射的温度。

由于黑洞霍金辐射的温度可用多种方法证明，所以上述讨论又反过来证明了克鲁斯卡时空（坐标系）确实处于绝对零度。

从第六章可以看出，绝对零度的克鲁斯卡坐标系可以接触史瓦西时空的内禀奇点（ $r=0$ ），但有限温度的史瓦西坐标系（有霍金辐射）则接触不到此内禀奇点。

此外，史瓦西黑洞在蒸发到最后时，其质量 M 越来越小，半径 $r=2M$ 也越来越小，温度

$$T = \frac{1}{8\pi K_B M} \quad (8.4)$$

越来越高，在最后消失的一瞬间， $M \rightarrow 0$ ， $r \rightarrow 0$ ， $T \rightarrow \infty$ 。有人研究认为，在最后一瞬间，史瓦西黑洞的奇点会在消失前裸露，但这时它的温度已是无穷大了。

值得注意的是，有一大类奇点（类时奇点，例如R-N黑洞的奇点，如图8-5中的曲折线所示）是类时线不可能抵达的，只有类光线（图中虚线）或趋近类光的类时线（图中实线）才能抵达。梁灿彬与盖若什等人首先证明了这一点。还可以证明，这种类时线在趋近奇点时，加速度

将趋于无穷大。依据第六章介绍的安鲁效应，沿这种类时线运动的观测者（或物体）在到达奇点时，温度将发散。

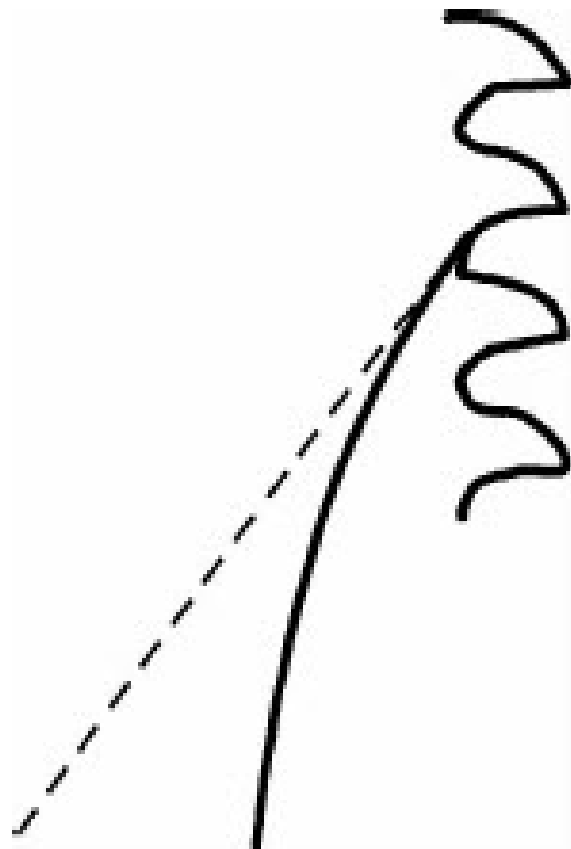


图8-5 趋近R-N黑洞奇点的火箭

用安鲁效应研究测地线的温度

我们还注意到，证明奇点定理所用的世界线都是测地线（类时、类光两种）。按照式（6.10）所示的安鲁效应，沿类时测地线（加速度为零）运动的观测者处于绝对零度。我们最近证明了类光测地线可以看作加速度为无穷大的类时线，按照安鲁效应，它对应温度发散的情况。

伦德勒曾经指出，平直时空中某种特殊的类光线，可以看作固有加速度为无穷大的类时线。图8-6为闵氏时空图。图中双曲线即伦德勒观测者的世界线，该观测者的固有加速度为有限值。研究表明，加速度越大，双曲线越趋近其渐近线。该渐近线在图中呈 45° 角，为类光线。当双曲线趋于其渐近线（类光线）时，沿此曲线运动的伦德勒观测者的固有加速度将趋于无穷大。所以，伦德勒得到上述结论。

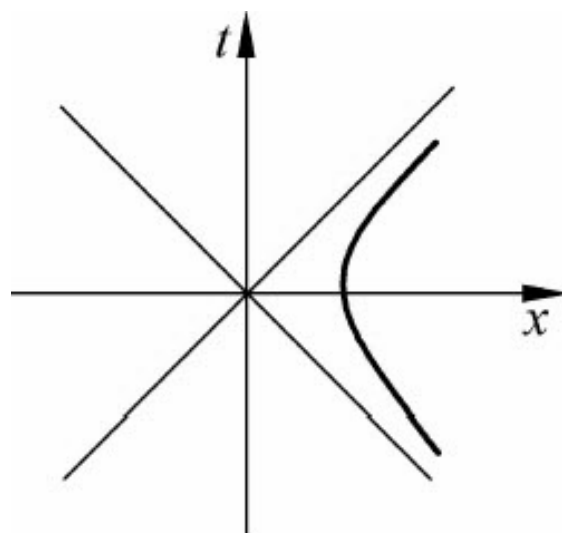


图8-6 伦德勒观测者的世界线

不过，这种特殊的类光线存在突变的拐点，相当于被镜子反射了一下，因而不是类光测地线。我们曾把伦德勒的这一结论，推广到弯曲时空中可以无限延伸的类光测地线。

下面我们将指出，这一结论对于存在共轭点的类光测地线也适用。这就是说，在奇点定理的证明中所使用的类光测地线（这种类光线会碰到奇点，不能无限延伸），确实可看作固有加速度为无穷大的类时线。

对类光测地线加速度的研究

众所周知，类光测地线不能用“固有时间”来描述，而只能用另一类仿射参量描述，因此不能直接对类光测地线定义加速度。我们证明的途径是，把类光测地线看作一族类时线汇的极限线。在类时线上可以严格定义加速度，然后让这族类时线趋近作为极限线的类光测地线，把这样得到的极限加速度定义为类光测地线的加速度。

关于存在共轭点的类光测地线，霍金等人曾证明一条定理：设 p, q 是光滑因果线 μ 上的两点，不存在连接 p, q 两点的光滑单参因果曲线（因果曲线指线上任二点有因果联系的曲线，即类光线和类时线）族 γ_μ （ $\gamma_0 = \mu$ ；当 $\mu > 0$ 时， γ_μ 类时）的充要条件是： γ_0 是一条类光测地线，且在 p, q 之间不存在与 p 共轭的点（图8-7）。

从这条定理可知，当 γ_0 上存在共轭于 p 的点 r 时，一定可以从 γ_0 微扰出因果曲线族 γ_μ 。除 γ_0 为类光测地线外， γ_μ （ $\mu > 0$ ）都是类时线。不难看出，

γ_0 是类时线汇 γ_μ 的极限线。图8-7中 ν_0^a 为 γ_0 的切矢， Z^a 为偏离矢量。

我们定义 γ_μ 上的加速度 A ，然后令 γ_μ 逼近 γ_0 ，发现加速度 $A \rightarrow +\infty$ 。

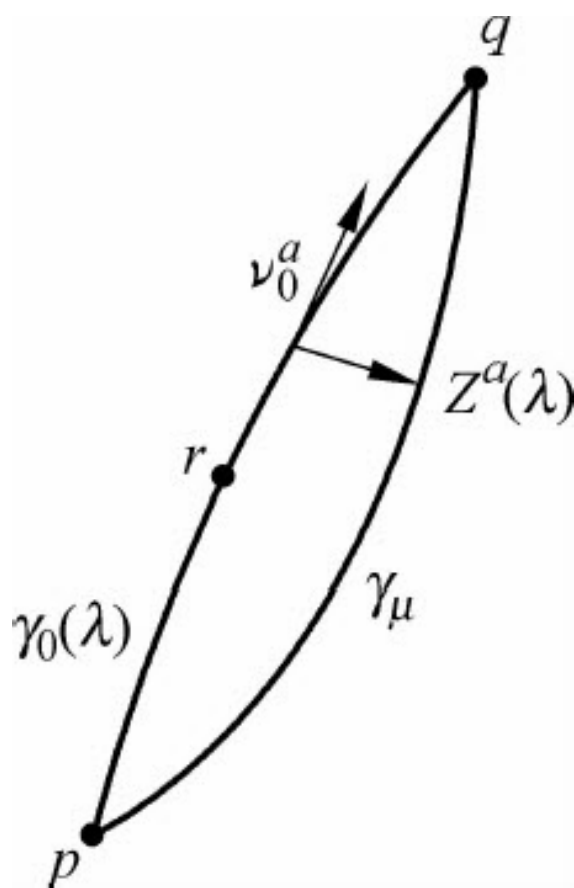


图8-7 类光线 γ_0 及微扰出的类时线 γ_μ

这样，我们就证明了 γ_0 的加速度 A 发散。这就是说，有共轭点的类光测地线 γ_0 ，可以看作固有加速度为无穷大的世界线。这是一个极具启发性的结果。

田贵花博士和我合写的相关论文已发表在美国的《数学物理》及欧洲出版的《经典与量子引力》等中外杂志上。

通常认为，自由光子作惯性运动，其固有加速度当然是零。现在我们看到一个惊人的相反结论，作惯性运动的光子的固有加速度居然是无穷大。这启示我们在光、惯性与时间的背后，存在重要的未知关系。光在相对论中已经处于核心地位，但从本节的讨论来看，我们对光的认识还非常不够，需要进一步深化。

奇点的证明与热力学第三定律冲突

大家注意，霍金和彭罗斯证明他们的奇点定理时，是用测地线做工具的。奇点定理表明，至少有一条光速或亚光速的测地线会在有限的时间（即测地线长度）内碰到奇点而断掉。所谓测地线，就是弯曲时空中的“直线”，沿测地线运动的物体，作的是惯性运动。

由安鲁效应可以推测，沿亚光速测地线运动的物体处在绝对零度。现在我们又证明，沿光线（类光测地线）运动的光子加速度为无穷大，依据安鲁效应可以把它们的温度视作无穷大。

对于通常的系统，热力学第三定律指出，不能通过有限次操作把系统的温度降低到绝对零度。因此，对于通常的热力学系统，它的温度定义在无穷大和绝对零度之间。研究表明，能量有限、能级数目也有限的系统，存在负温度，它的温度定义在负零度（ -0K ）与正零度（ $+0\text{K}$ ）之间。这种系统称为负温系统。

在负温系统中，热力学第三定律表为，既不能通过有限次操作把系统温度降到正零度，也不能通过有限次操作把系统温度升高到负零度。对于通常的热力学系统，第三定律只说了下限绝对零度不可抵达，未涉及上限，但上限“无穷大”显然也不可能抵达。

实际上，两类系统的温度都定义

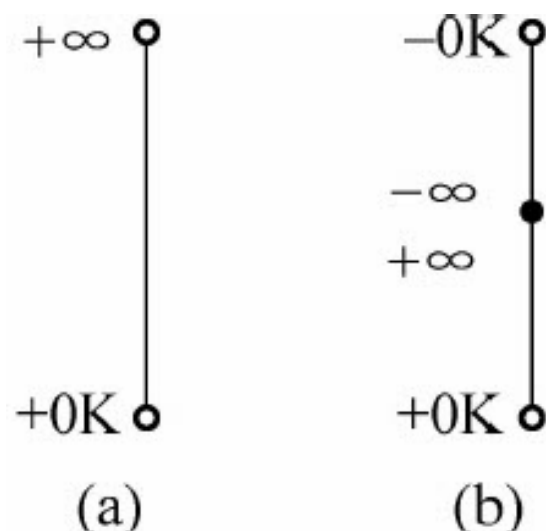


图8-8 两种热力学系统的温度区间

(a) 常温系统； (b) 负温系统

在开区间上，温度的上限（无穷大或负温度系统的 -0K ）与下限（绝对零度）均不能通过有限次操作达到，所以可以一般性地把第三定律推广到包括上限的情况。因此，我们认为，奇点定理是在违背热力学第三定律的情况下证明的。

上述情况提示我们，奇点的出现或达到奇点，往往伴随温度异常（绝对零度或无穷大）的情况出现。奇点的出现是违背第三定律造成的，也就是说，热力学第三定律可能会不容许时间有开始和终结。

前面谈到，许多人认为“宇宙监督”就是第三定律，第三定律会阻止奇点裸露。现在看来第三定律不仅会阻止奇点裸露，而且很可能会干脆阻止奇点存在。

彭罗斯和霍金把奇点看作时间的开始或终结，在认识上是一个重大的进展。奇点定理说，时间一定有开始或终结。现在我们看到，奇点定理与热力学第三定律冲突，因此我们推测第三定律将保证不存在奇点，保证时间没有开始和终结，因此时间应该是无限的，无始无终的。

我们对这些问题的有关科研论述，刊登在《中国科学》等杂志，以及作者的科研专著中。

第九章 时间测量的疑难与探索

法国数学家庞加莱认为“时间必须变成可测量的东西，不能被测量的东西不能成为科学的对象”。

然而，时间测量问题要远比人们想象的复杂得多。主要困难还不是测量的技术，而是测量的理论依据。下面我们就来叙述和讨论这一问题，读者会从中悟出深刻的哲理。

用周期运动度量时间

作为“绵延”的时间，有“持续性”和“流逝性”双重性质。

“流逝性”如何度量，在自然科学中尚无公认的方法。虽然普列高津等人有不少重要的探讨，虽然可以考虑用“不可逆熵”的产生率来进行度量，但离形成真正的科学定义尚有不小的距离。

“持续性”的度量则从古就有。从人类文明诞生的时候起，古人就用各种周期运动来度量时间，例如地球自转、公转引起的星空的周期变化，月亮的周期性圆缺，日影高度的周期变化，以至后来的单摆周期运动等。

近现代的各种钟表，都是用周期运动来度量时间，即度量时间的“持续性质”。以后我们谈到时间的测量或度量，都是专指对时间“持续性”的测度。对于目前尚不清楚的“流逝性”的度量，我们暂不讨论。

综上所述，人类长期以来是通过运动，主要是周期运动来度量时间。亚里士多德就明确说过：

“时间是运动的计数。”

“时间是运动和运动持续量的尺度。”

似乎人们只要造出统一的、高质量的按某种科学规律的周期运动来计量的钟，再把这些钟分置到世界各地，度量时间的问题就彻底解决了。然而，深入的哲学思考却使这似乎不成问题的问题成了大问题。

度量时间需要先做“约定”

牛顿在阐述自己的绝对时空观时就明确提出，绝对时间是客观存在的，均匀流逝的，但却是不可测量的。

通常悉知的可测量的时间，是对运动延续的度量，只不过是绝对时间的代替物，是一种相对时间，表观时间。牛顿清醒地认识到：“可能并不存在一种运动可以用来准确地测量（表观）时间”，“所有的运动可能都是加速或减速的”。

与牛顿同时代的著名的哲学家洛克指出：时间是对绵延长度的度量。绵延只能用周期运动作单位进行度量，然而“绵延中任何两部分，我们都不能确知是相等的”。我们只能假定，周期运动的每一个周期都是相等的，才能对时间进行度量。

洛克的意思是，我们无法知道周期运动的每一个周期都是相等的。造成这种局面的原因在于时间与空间不同。在洛克那个年代，人们认为空间是永存的，尺子也是不变的。我们只要选择一把尺子，把它的长度作为标准长度，就可以反复测量同一个长度。

因此人们认为空间的度量不成问题。然而时间却不同，时间具有一去不复返的性质，“过去”已经不存在了，“未来”还没有来临。

我们无法把周期运动的一个周期移动到另一个周期去进行比较，因此，各种周期的相等只有靠“约定”（即“规定”）。只有“约定”了某种周期运动的各个周期相等，我们才能测量时间。

用运动定律度量时间——“好钟”

18世纪著名的数学家欧拉提出一个新思路：用运动定律来确认周期运动的各个周期相等。他在《时间和空间的沉思》一书中写道：如果以某个给定的循环过程为单位时间，而发现惯性定律成立的话，这个过程就是周期的。即每次循环都经历相同的时间，或者说各个时间周期相等。

欧拉所处的时代，相对论尚未诞生，大家认为空间的测量不成问题。尺子可以来回移动，反复测量。

在标度好了距离的空间，使用尚需考察的钟表来计量时间，如果把这样测得的时间与空间距离相配，在任何相等的时间段内，不受外力的质点都走过相等的距离，那么就可以认为这个钟是“好钟”，它的各个周期是相等的。

欧拉的思想后来被进一步发展，认为“好钟”可以这样定义：用它标度的时间，应该保证牛顿三定律成立，麦克斯韦电磁理论成立，能量动量守恒定律成立，等等。一句话，应该保证物理规律简单！

然而，怎么严格定义“物理规律简单”呢？我们如何确认现在已知的物理规律表达式已经是最简单的了？这个问题很难回答。

相对论诞生之后，人们逐渐认识到不仅时间的测量有问题，空间的测量也存在问题。一根尺子在不同的地方，长度是否一样？尺子在移动过程中长度会不会改变？都成了需要深思的问题。

庞加莱的推测——“约定光速”

哲学家们在探索“时间本质”上所发挥的激情，使时间测量问题更加混乱。不少哲学家认为时间与空间不同，时间应该属于精神世界。有的哲学家干脆认为时间的度量只能靠“直觉”。然而什么是“直觉”，也很难说清楚，似乎只能意会，不能言传。

哲学家的聪明才智虽然给了科学工作者很多启示，但上述把时间归入“精神世界”，把时间度量归入“直觉”的看法，似乎无助于自然科学工作者对时间性质的研究。

针对上述导致“混乱”的观点，庞加莱在相对论诞生前夜（1900年前后）发表了一些重要看法。庞加莱认为时间的测量分为两个问题，一个是如何确定“异地时钟”的“同时”，另一个是如何确定“相继时间段”的“相等”。他认为这两个问题的解决不能靠“直觉”，而应靠“约定”。

那么，约定什么呢？1898年，庞加莱在《时间的测量》一文中猜测，应该把“光速各向同性而且是一个常数”作为一条公理（即约定）。他讨论了用交换光信号来确定两地时间“同时”的问题。

1905年，他又在《科学的价值》一书中再次强调了他对“约定”光的传播性质的观点：“光具有不变的速度，尤其是，光速在所有方向都是相同的。这是一个公设。没有这个公设，便不能试图测量光速。”

正是这些见解，引导爱因斯坦走上了发现相对论的道路。

爱因斯坦遇到难解之谜

科学史的研究表明，在相对论的第一篇论文发表之前很久，爱因斯坦就已认识到“相对性原理”和“麦克斯韦电磁理论”是应该坚持的基本原理。他也已认识到这将导致电磁理论与参考系无关，以及由此引起的光速与参考系无关的结论，即所谓“光速不变性”。也就是说爱因斯坦已经抓住了相对论的基础。

那么他为什么一直没有建立起相对论呢？爱因斯坦1922年在日本京都的一次即兴演讲道出了其中的原委。他回忆了大约在1905年5月与朋友贝索的一次讨论，当时爱因斯坦正被一个问题卡住。这个问题就是“光速不变性”（即光速与观测者相对于光源的运动速度无关），似乎与力学中的速度叠加法则（平行四边形法则）相矛盾。这个难题爱因斯坦思考了几乎一年，然而毫无结果。他觉得“这真是个难解之谜”。

爱因斯坦在京都演讲中回忆道：“这时，伯尔尼的一个朋友（贝索）意外地帮助了我。那是一个明媚的日子，我去访问他，与他进行了如下的谈话：‘最近我有个难以理解的问题，所以今天我把问题带到这里来想跟你讨论。’我们谈了很多，我突然明白了。第二天我又去看他，开口就说：‘谢谢你，我已经完全解决了这个问题。’我解决的实际上就是时间概念，也就是说，时间不可能被绝对地定义，时间和信号速度之间存在着不可分离的联系。”

庞加莱对爱因斯坦的启发

在专利局工作时，爱因斯坦和他的朋友们组织了一个自称为“奥林匹亚科学院”的自由读书与讨论的“俱乐部”。他们曾一起读过并讨论过庞加莱的文章。看来这次与贝索的谈话使爱因斯坦回忆起了庞加莱关于时间与光速关系的论点，这给了他重要的启示，解决了那个被卡住的问题。

几周后爱因斯坦关于相对论的第一篇论文《论运动物体的电动力学》就投给了杂志社，文章后面，爱因斯坦向贝索致谢，“感谢他提出的一些有价值的建议。”

需要指出的是，爱因斯坦从来没有谈到过庞加莱在这方面的启发，也没有强调过他在“奥林匹亚科学院”的活动中阅读、讨论过庞加莱的著作。他只是一再强调他读过马赫的书籍，一再强调马赫对他的影响。然而，他在“奥林匹亚科学院”中的伙伴们后来回忆过他们一起阅读庞加莱作品的情况，谈到阅读庞加莱文章时的激动与兴奋，这种情绪一直感染了他们几个月。

爱因斯坦受到过庞加莱的启发应该是显然的，但是他为什么不谈这一点呢？这可能是由于他与庞加莱关系不太好所致。

爱因斯坦与庞加莱只见过一次面，那是在第一次索尔维会议上。当时庞加莱已是举世闻名的数学大师，爱因斯坦只是一个初出茅庐的新秀。爱因斯坦本指望庞加莱会支持他的相对论，但会后他十分沮丧，对他的朋友说：“庞加莱根本不懂相对论。”

后来，苏黎世大学想聘爱因斯坦去当教授，征求庞加莱的意见，庞

加莱写了如下回复：“爱因斯坦先生是我所知道的最有创造思想的人物之一，尽管他还很年轻，但已经在当代第一流科学家中享有崇高的地位。……不过，我想说，并不是他的所有期待都能在实验可能的时候经得住检验。相反，因为他在不同方向上摸索，我们应该想到他所走的路，大多数都是死胡同；不过，我们同时也应该希望，他所指出的方向中会有一个是正确的，这就足够了。”

总之，庞加莱对爱因斯坦评价不高，从来没有表示过赞同相对论。而且不久之后，庞加莱就去世了，没有来得及改变自己的看法。

下面，我们将清楚地看到，庞加莱“约定”光速的观点，对爱因斯坦建立相对论的影响。

爱因斯坦的突破：“约定光速”来定义“同时”

“异地时钟的校准”和“相继时间段（绵延）的测量”是时间研究中的重大问题。庞加莱认为这两个问题相互关联，而且只有通过“约定”才能加以解决。他推测通过“约定”真空中光速的各向同性有可能解决上述问题。

爱因斯坦在他的相对论中正是用“约定光速”的方式，定义了异地事件的“同时”。由于物理学是一门实验的科学、测量的科学，有关时间度量的任何约定，都必须使定义在测量上有可操作性。

为此，爱因斯坦建议“约定”真空中的光速均匀各向同性，而且是一个常数。我们从这一约定中，不难看到庞加莱猜想的影子，爱因斯坦把庞加莱的猜想，具体实践到了物理理论中。

在相对论的开创性论文《论运动物体的电动力学》中，爱因斯坦给出了“同时性的定义”。他写道：

“如果在空间的A点有一个钟，在A点的观察者只要在事件发生的同时，记下指针的位置，就能确定A点最邻近的事件的时间值。若在空间的另一点B也有一个钟，此钟在一切方面都与A钟类似，那么在B点的观察者，就能测定B点最邻近处的事件的时间值。

但是若无其他假设，就不能把B处的事件同A处的事件之间的时间关系进行比较。到目前为止我们只定义了‘A时间’和‘B时间’，还没有定义A和B的公共‘时间’。”

爱因斯坦接着写道：

“除非我们用定义规定，光从A走到B所需的‘时间’等于它从B走到A所需的‘时间’，否则公共‘时间’就完全不能确定。

现在令一束光线于‘A时刻’ t_A 从A射向B，于‘B时刻’ t_B 又从B被反射回A，于‘A时刻’ t'_A 再回到A（图9-1）。

按照定义，两钟同步的条件是

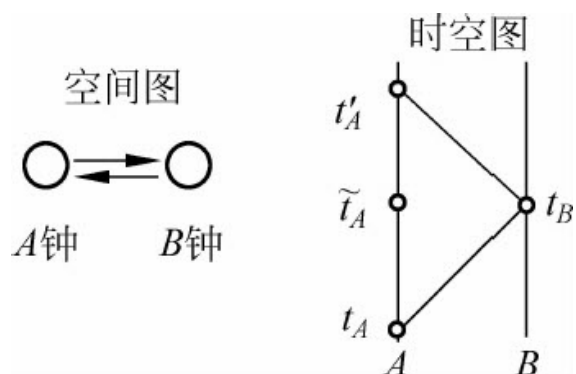


图9-1 惯性系中异地时钟的校准（空间图与时空图）

$$t_B - t_A = t'_A - t_B \quad (9.1)$$

我们假定，同步性的这个定义是无矛盾的，能适用于任何数目的点，并且下列关系总是成立的：

（1）假如B处的钟与A处的钟同步，则A处的钟与B处的钟也同步。

（2）假如A处的钟与B及C处的钟同步，则B、C两处的钟彼此也同步。

这样，借助于某些假想的物理实验，我们解决了如何理解位于不同地点的同步静止钟这个问题，并且显然得到了‘同时’或‘同步’的定义，以及‘时间’的定义。”

爱因斯坦又写道：“根据经验，我们进一步假设，令

$$\frac{2AB}{t'_A - t_A} = c \quad (9.2)$$

是个普适恒量，即在真空中的光速。”

式（9.1）可改写为

$$\frac{t_A + t'_A}{2} = t_B \quad (9.3)$$

爱因斯坦就把A钟的时刻

$$\tilde{t}_A = \frac{t_A + t'_A}{2} \quad (9.4)$$

定义为与B钟的 t_B 同时的时刻。在平直时空的惯性系中，爱因斯坦用这种方法定义了异地静止时钟的“同时”。在操作过程中，他上面提到的几点假设都没有出现矛盾。

朗道的发现：“同时的传递性”需要条件

然而，后来的研究却出人意料，人们惊讶地发现，如果在平直时空中采用非惯性系，或在弯曲时空中采用任意的曲线坐标系，则爱因斯坦的假设（2）不一定成立。研究发现，只有在时轴正交系（时间轴垂直于三个空间轴）中，“同时”才具有传递性（即假设（2）成立），才能在时空中建立“同时面”，定义统一的时间，使各点的钟保持“同时”和“同步”。

下面我们介绍一下朗道等人关于“同时”传递性的讨论，即对爱因斯坦所提的假设（2）在什么条件下成立的讨论。

爱因斯坦的假设（2）说，在任何一个惯性系中固定三个钟，如果用上面的方法把A、B两个钟对好，再把B、C两个钟对好，那么A、C两个钟就自然对好了。这表示“同时”这个概念具有传递性，全时空可以定义统一的时间。此结论与人们的常识一致，因此没有引起更多的兴趣。

然而朗道等人发现，在弯曲时空中（以及平直时空的非惯性系，例如在转动的圆盘上）却未必一定能做到这一点。在弯曲时空的一个参考系中，如果放置三个固定钟，同样可以用爱因斯坦建议的方法来“对钟”。可以把A、B两个钟对好，再把B、C两个钟对好，但是，这时A、C两个钟却往往对不上（图9-2）。

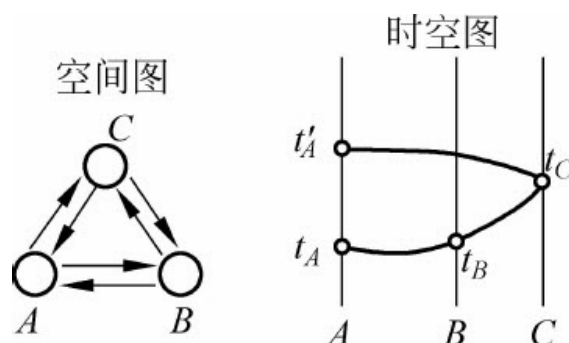


图9-2 “同时”的传递性（空间图与时空图）

这表示“同时”这个概念不具有传递性，不能在全时空定义统一的时间。朗道等人证明，只有在时间轴与三个空间坐标轴都垂直的情况下，即所谓“时轴正交”的情况下，A、C两个钟才能自然对好，“同时”才具有传递性，才能在全时空定义统一的时间。

朗道等人给出的时轴正交条件是

$$g_{0i} = 0, \quad i = 1, 2, 3 \quad (9.5)$$

我们在第五章中指出，度规的 g_{0i} 分量为零，表示时间轴与空间轴 x^i 垂直。“时轴正交”是指时间轴与三个空间轴都垂直，即 g_{01} 、 g_{02} 、 g_{03} 均为零的情况。

狭义相对论通常使用以直角坐标描述的惯性系，时空的四个坐标轴均两两垂直，当然时轴正交，“同时”自然具有传递性。所以，爱因斯坦提出的假设（2），在狭义相对论的研究中没有出现问题。

但是，广义相对论中的坐标系是任意的曲线坐标系，一般都不会时轴正交。不过，在大多数时空中，都可以找到一个时轴正交的坐标系来做参考系，在这个参考系中“同时”具有传递性，可以定义统一的时间。然而，如果在一个时空中找不到这样的时轴正交系，那么就不可能在此时空中定义统一的时间。

因此，在相对论中“同时”这个概念不是在任何参考系中都具有传递性的。也就是说，不是在任何参考系中，都能使静止在各点的钟“同时”或“同步”，这就是说，不是在任何空间中，都能建立统一的时间。广义相对论明确地给出了“同时”具有传递性的条件-时轴正交，就是时间轴必须与三个空间轴都垂直，三个空间轴之间是否相互垂直，没有关

系。

突发奇想

当我在刘辽先生的“广义相对论”课上学到朗道的“对钟理论”时，感到十分震撼：自然界中居然存在这样的怪事！A钟与B钟对好，B钟再与C钟对好，C钟却与A钟对不上。这太令人匪夷所思了！

在震惊之余，我不禁去想，自然界中还有没有类似的情况呢？我突然想到了热力学第零定律。第零定律告诉我们“热平衡具有传递性”。三个物体，如果A与B达到热平衡，B与C达到热平衡，A与C就自然达到了热平衡。

我从小就喜欢胡思乱想，这时突发奇想：“同时具有传递性”是否和热力学第零定律等价呢？当时我主要在跟随刘辽先生研究黑洞和引力波，但这个念头不断出现在我的脑海中，我终于决定不妨试一试，看看能否证明二者等价。

要进行这一论证，首先要找到合适的数学物理工具，这一工具必须能在“时间”和“温度”之间建立桥梁。于是在主要工作之余，我开始寻找这一桥梁。

不久我就选择了“温度格林函数”，即松原函数。我不是专门研究统计物理和热力学的，但在研究黑洞热辐射时，接触到过松原函数，对它有一点了解。在松原函数的参量中有温度，还有与时间有关的量。

自此我就投入了对此问题的研究，很快发现，在松原函数中，与温度相关联的不是“时刻”，而是“时间周期”（实际上是虚时间的周期）。我觉得，这样看来，如果“时间”与“温度”真的有关联，那么与热力学第零定律等价的大概不会是“时刻”的传递性，而可能是“时间周期的传递

性”，也即“钟速快慢”的传递性。

对钟的新等级：钟速同步传递性

于是，我认为应该引入一个新的概念：“钟速同步的传递性”，并找出其成立的条件。这就是说，我不要求A、B、C三个钟的时刻对好，只要求它们的快慢对好。也就是说，不要求三个钟的指针总是指同一时刻，只要求三个钟的钟速能够总“同步”，能够总走得一样快。显然，这应该是一个比朗道的“对钟”等级，要宽松的等级。

所谓钟的快慢，就像图9-3中线段的长短，如果这一段 $t_{A2}-t_{A1}$ 等于 $t_{B2}-t_{B1}$ ， $t_{B2}-t_{B1}$ 等于 $t_{C2}-t_{C1}$ ，是不是转回来的这一段 $t'_{A2}-t'_{A1}$ 和原来这一段 $t_{A2}-t_{A1}$ 是一样长的。如果是一样长的，就是“钟速同步具有传递性”。

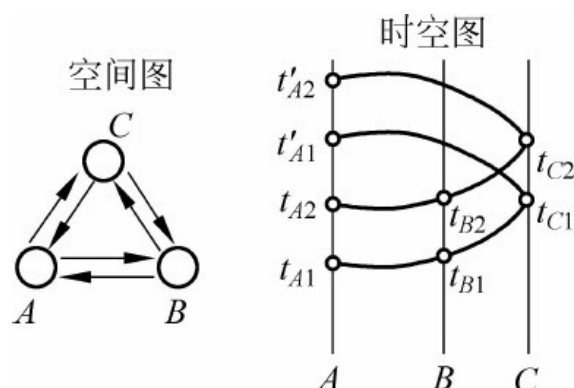


图9-3 “钟速同步”的传递性（时空图）

我按照爱因斯坦和朗道的对钟方案，沿着它们的思路，研究了“钟速同步的传递性”，给出了这一传递性成立的条件

$$\frac{1}{x^0} \left(\frac{g_{03}}{g_{00}} \right) = 0 \quad (9.6)$$

式中 $x^0=ct$ 。

显然，此条件弱于朗道给出的“同时具有传递性”的条件式（9.5）。式（9.5）成立的时空，式（9.6）一定成立，反之则不一定。这就是说，“同时具有传递性”的时空，“钟速同步”一定具有传递性，反之则不一定。

直观一点说，就是如果静止于空间各点的钟，总是指着相同的“时刻”，你的钟3点，他的钟也3点，大家都是3点；你的钟4点，他的钟也4点，大家都是4点。这种情况下，各个钟的快慢必定相同。

反过来，各点的钟快慢都相同，并不一定各钟的指针都指着相同的时刻，只表示它们指针的时刻一定总保持相同的差值。

“钟速同步传递性”等价于热力学第零定律

有了“钟速同步传递性”这一概念及其成立条件式(9.6)之后,我就利用松原函数进一步探讨了“钟速同步传递性”与“热平衡传递性”之间的关系。研究表明,二者是等价的,“钟速同步”具有传递性的时空,热力学第零定律一定成立,反之亦然。

我们看到,只有在热力学第零定律成立的参考系中,才能在空间各点建立“速率”相同的时钟,即保证位于空间各点的静止钟能走得一样快,总是保持“同步”。

不过,这还不能保证静止于空间各点的钟做到“同时”。那是因为保证“同时”具有传递性的条件式(9.5)比式(9.6)更苛刻。式(9.6)只是式(9.5)的必要条件而不是充分条件。

爱因斯坦当年的论文没有区分钟的“同时”和“同步”这两个概念,我们的研究表明,这两个概念实际上有区别。做到“同步”的钟,并不一定指着同一时刻;但做到永远保持“同时”的钟,一定保持同步。

式(9.5)成立的参考系,时轴正交。在其中“同时具有传递性”,因而可以在整个空间定义统一的时间,在几何上画出同时面。这样的时空,式(9.6)必定成立,“钟速同步”一定具有传递性,热力学第零定律一定成立。

反过来,热力学第零定律成立的时空,一定能在空间各点建立相同的温标,静止于空间各点的钟也可以调整到“同步”,有统一的钟速,但还不一定能有统一的时刻,即不一定能建立同时面。

从1985年开始，我把这一工作写成论文，先后发表在《中国科学》等国内外杂志上，以及科研著作中。从1988年开始，又在我讲授的“广义相对论”课中讲到了上述工作，并把式（9.6）写入了自己编写的讲义。这套讲义在北京师范大学使用多年，后来又在清华大学讲课时使用过，最后由清华大学出版社以《广义相对论基础》的书名正式出版，不过那已是2010年的事情了。

1996年前后，北京师范大学的一位研究生高思杰在听了我的课后，认为我搞的这个研究工作有道理，就做了进一步的研究和发展，并与中科院数学所的邝志全研究员、北京师范大学的梁灿彬教授合作发表了一篇论文，确认了“钟速同步”是一个新的对钟等级。

现在，关于“钟速同步传递性”的研究成果，已出现在梁灿彬先生和周彬博士合写的《微分几何与广义相对论》一书中。

不过，他们在公开发表的文章和书籍中只是对“钟速同步传递性”发表了意见，未对此传递性是否与第零定律有关发表意见。

此后，我又曾多次回到上述问题的研究，并用“普朗克黑体谱”和“测不准关系”对第零定律与“钟速同步传递性”的关系进行了新的论证。

总之，我们的研究表明，热力学第零定律等价于“钟速同步”具有传递性，或者说，第零定律保证时空中一定存在一个“钟速同步具有传递性”的参考系。也就是说，第零定律是保证“同时”具有传递性的必要条件，或者说，保证时空中一定存在一个时轴正交系（因而在空间各点可以定义统一的时间）的必要条件。

所以，“同时”这个概念，并不是在任何空间中都存在的，只有在热

力学第零定律成立的空间中，它才可能存在。

时间这个概念，并不是在任何时空中都可以统一定义的，只有在热力学第零定律成立的空间中，才可能对空间各点定义统一的时间。粗略地说，热力学第零定律告诉我们，“同时”是可以定义的，或者说，时间是可以在空间各点统一定义的。

在热力学中，第零定律的作用，是保证温度可以定义。我们惊讶地发现，它也保证时间可以定义。这是一个有趣的结论。

那么，为什么以往没有人注意到这个问题呢？这是因为物理学的各个领域，除去广义相对论外，都不考虑时空弯曲，都只在狭义相对论和惯性系的范围内讨论问题，所用的坐标系都是时轴正交系。热力学定律也不例外。因此以往很少有人注意非惯性系和时空弯曲对热力学的影响。

有关“钟速同步传递性”等价于热力学第零定律的论述，后来收入了我所写的科研专著和教材中。

绵延的相等：“时间段”相等的定义

庞加莱认为“异地时钟的同时”和任何一个时钟的“相继时间段（绵延）的相等”是时间测量中的两个重大问题。他认为这两个问题相互关联，而且只有通过“约定”才能加以解决。他推测通过“约定”真空中光速的各向同性有可能一起解决上述两个问题。

我非常赞同庞加莱的设想，即对光速传播性质的约定，应该是全部时间测量的基础，不仅“异地时钟的同时”，而且同一时钟“相继时间段的相等”都应该能通过“约定光速”来加以解决。

然而，爱因斯坦和朗道虽然用这一约定讨论了异地时钟同时的问题，但他们都没有讨论同一时钟的“相继时间段相等”应如何定义的问题。

前面提到过，对于“相继时间段的相等”，物理学中使用的是所谓“好钟”的概念。该定义源于欧拉的思想：如果以某个给定的循环过程为单位时间，而发现牛顿第一定律成立，这个过程就是周期的。

现代广义相对论中采用了欧拉的思想，认为一个“好钟”所走的时间，应该保证惯性定律成立，自由质点在局部时空区的时空轨迹应该是直线。这一思想还被进一步发展：“好钟”的计时应保证物理规律简单，例如牛顿定律成立，麦克斯韦电磁理论成立，能量守恒定律成立，等等。

因此，现代物理学中时间与空间的测量，建立在两个约定的基础上。一个是约定光速均匀各向同性，而且是一个常数 c ，另一个是约定存在保证物理规律简单的“好钟”。在这两个约定的基础上，可以定义不

同空间点的钟“同时”或“同步”，还可以定义绵延的相等，并进一步用光速乘时间，定义空间距离。

但是，什么叫物理规律简单是个很难说清楚的问题。要求在局部惯性系中惯性定律成立，则必须事先定义“标准尺”，有了正确的空间距离，才能通过惯性定律验证钟的好坏。

然而，在广义相对论和现代物理学中，空间距离是用“约定的光速”乘上光传播这段距离的时间来定义的。我们看到，定义“时间”要用到“距离”；定义“距离”又要用到“时间”，这里面存在逻辑循环。因此“好钟”的定义有它的内在矛盾。

前面还谈到，我们建议了一个新的对钟等级：我们沿用爱因斯坦-朗道采用的对真空中光传播性质，即“光速各向同性而且是一个常数”的约定，给出了钟速同步具有传递性的条件。满足这一条件的时空，虽然不一定有“同时面”，但在全时空可以有统一的“钟速”。在相关论文中，我们强调的是异地时钟的钟速同步。

下面将指出，通过我们给出的钟速同步具有传递性的条件，不仅可以解决“异地时钟钟速同步”的定义问题，而且可以解决同一时钟“相继时间段相等”的定义问题。我们的这一研究结果，补充并发展了庞加莱和爱因斯坦等人关于时间测量的理论。

“约定光速”可以定义“相等的时间段”

大约在2005年左右，在参与“纪念相对论发表100周年”的宣传活动时，我重新阅读了爱因斯坦创立狭义相对论的第一篇论文《论运动物体的电动力学》及庞加莱的《科学与假设》、《科学的价值》等书籍，再次注意到他们对时间测量问题的论述，特别是物理学中对解决“异地时钟同时”和“相继时间段相等”的不同做法。

我想，是否能像庞加莱预期的那样，在“约定光速”的前提下，像解决“异地时钟同时”一样，来解决“相继时间段相等”的问题呢？

这时我回忆起当年自己提出“钟速同步传递性”这一概念时，有人问我，既然不同地点的钟可能指着不同的时刻，你谈的同步的“钟速”是哪个时刻的呢？我思考后回答说是“任何时刻”，我指的“钟速同步”，是说这些钟任何时刻都同步。对方大概觉得此答复还可以，就不再问了。

想到此，我沿着这一思路继续琢磨，既然“任何时刻”钟速都同步，如果把“对钟”操作持续一圈，“对”回到原来的钟，岂不就是同一时钟的“相继时间段相等”吗？

请大家看图9-3， $(t_{A2} - t_{A1})$ 和 $(t'_{A2} - t'_{A1})$ 不就是A钟在不同时刻的两个“时间段”吗？在此图中，把时间段 $t_{A2} - t_{A1}$ 与B、C两处的钟速（表现为时间段的长短）相继校准，就是使得 $t_{B2} - t_{B1} = t_{A2} - t_{A1}$ ， $t_{C2} - t_{C1} = t_{B2} - t_{B1}$ 。再把C钟的钟速与A钟的钟速校准，则得到时间段 $t'_{A2} - t'_{A1} = t_{C2} - t_{C1}$ 。这时，原来的 t_{A1} ，没有对回到原处，而到达了 t'_{A1} ， t_{A2} 也未对回到原处，而到达了 t'_{A2} ，但满足了 $t'_{A2} - t'_{A1} = t_{A2} - t_{A1}$ ，这两个

时间段相等了。

如果我们调节 $t_{A2}-t_{A1}$ 的长短，使其经B、C二处再对回A处时， t'_{A1} 恰好与 t_{A2} 重合（图9-4），则线段 $t_{A2}-t_{A1}$ 与 $t'_{A2}-t'_{A1}$ 恰好连接，于是我们得到了相等的“相继时间段”。这就正好解决了定义“相继时间段相等”的困难。

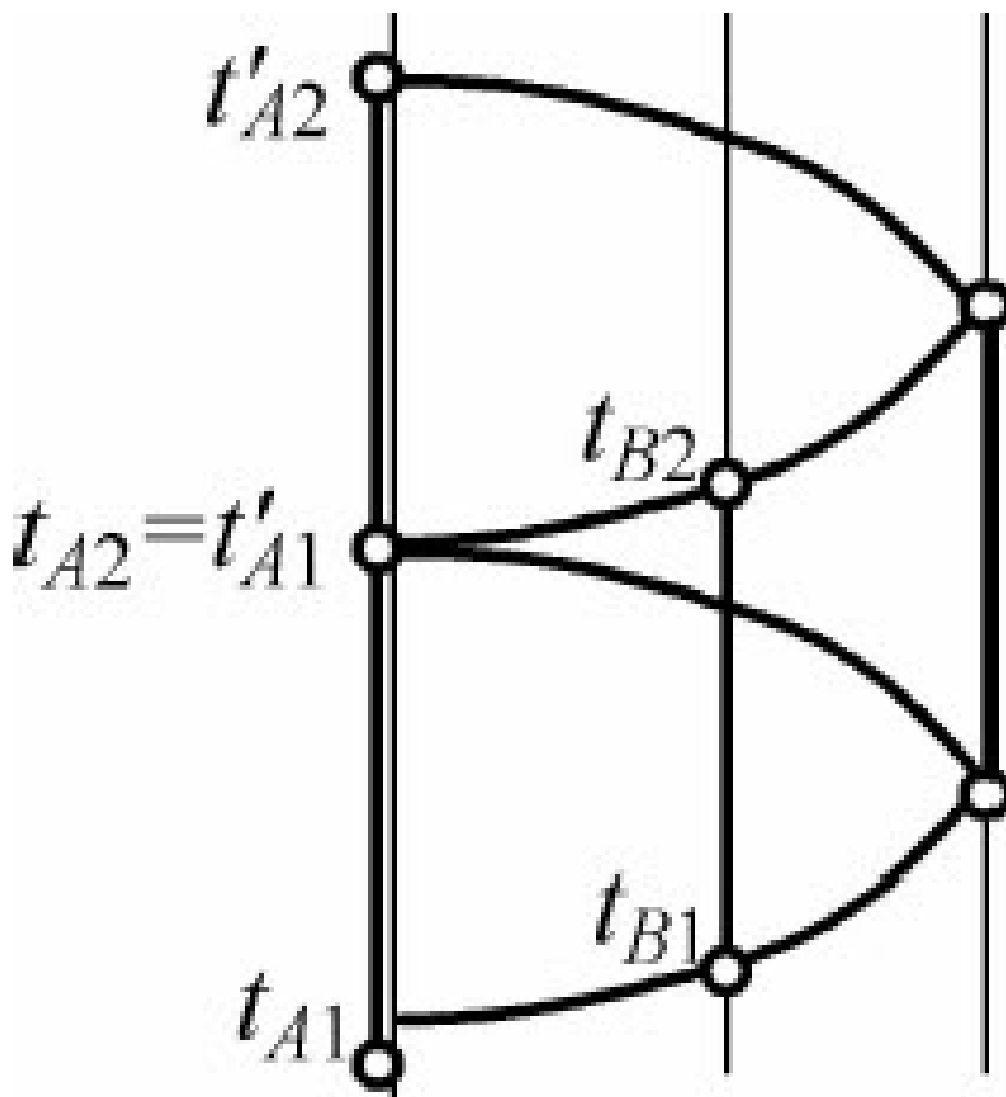


图9-4 相继时间段的相等（I）

事实上，操作可以更简化，只需要对一个时刻即可。如图9-5所

示，让A钟的 t_{A1} 经过与B、C钟的时刻对好，回到A钟的时刻为 t_{A2} 。然后把 t_{A2} 再做一次与B、C钟校对，再次对回A钟的时刻为 t_{A3} ，显然 $t_{A3} - t_{A2}$ ，即图9-4中的 $t'_{A2} - t'_{A1}$ ，也就是说“钟速同步的传递性”可以保证 $t_{A3} - t_{A2} = t_{A2} - t_{A1}$ ，也就保证了同一时钟“相继时间段”的相等。

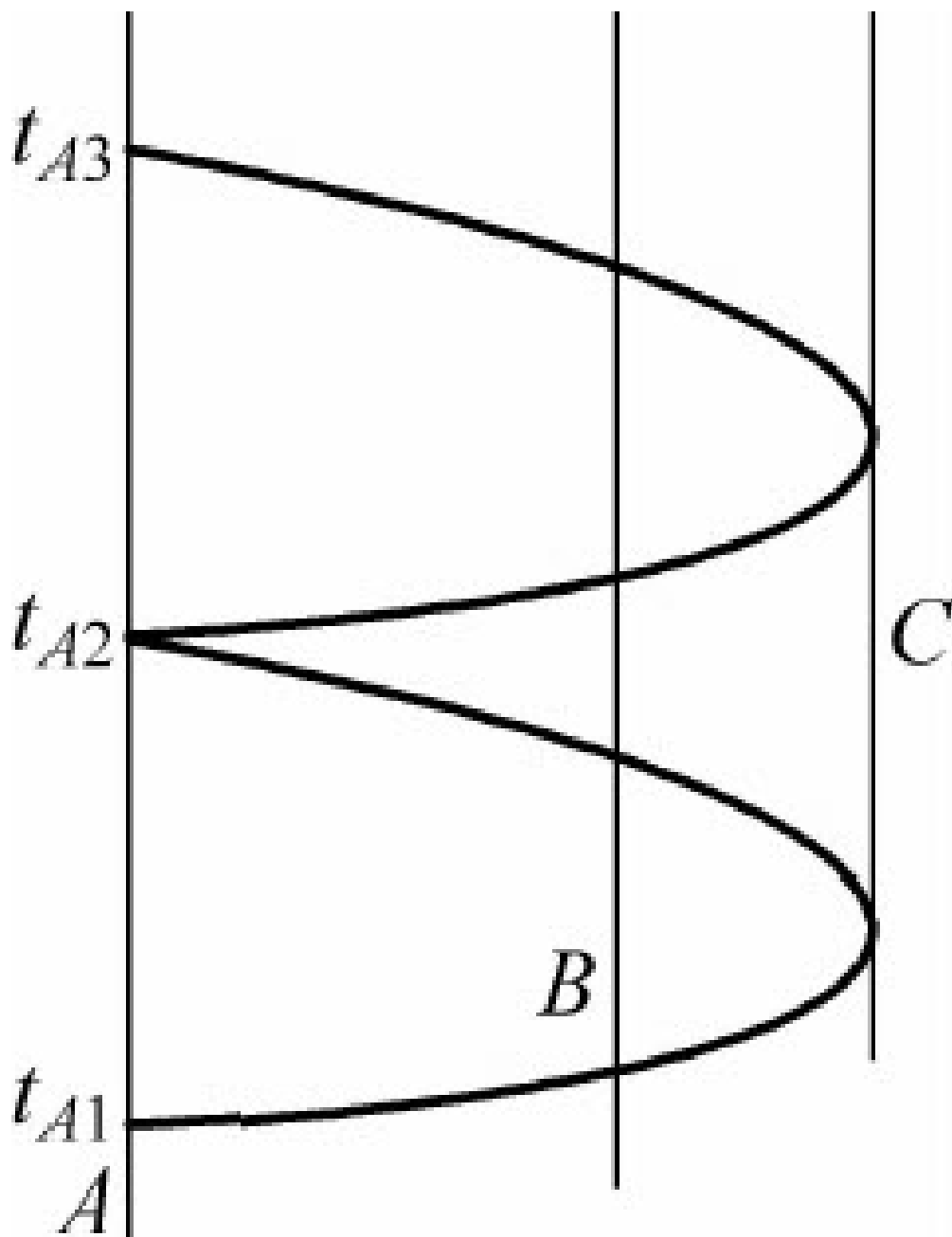


图9-5 相继时间段的相等（II）

我们看到，通过我们自己给出的调整“钟速同步”的方案，不仅可以用上述对光速的“约定”，定义异地时钟“钟速”的“同步”，而且可以定义任一指定坐标钟“相继时间段（又称绵延）”的相等。

我们的上述研究工作，不仅回答了庞加莱提出的时间测量的第一个问题（即异地时钟同步的定义），而且回答了他提出的第二个问题（即相继时间段相等的定义）。这使得我们能够在全时空定义统一的时间（即统一的坐标钟钟速）。因此，现代物理学中关于“好钟”的约定是不需要的。

此前我们的研究曾指出，热力学第零定律保证钟速同步具有传递性。现在我们进一步看到，热力学第零定律保证，不仅可以在全空间，而且可以在全时空，定义统一的时间。

有关内容我们发表在《中国物理快报》上，并收入进我们的几本科研专著。

“约定光速”等价于“约定时空的对称性”

不过，在“好钟”的理论中也还存在合理的因素。即要求能量守恒定律成立是个好的想法，它反映了时间流逝的均匀性。

事实上，对真空中光速的约定，就是对时空对称性的约定。要求真空中的光速点点均匀，相当于要求时间和空间具有均匀性，而时间和空间的均匀性又分别对应着物理学中的能量守恒和动量守恒。要求真空中的光速各向同性，相当于要求空间各向同性，它对应着物理学中的角动量守恒。要求光速不变原理成立，则相当于要求时空存在所谓布斯特（Boost）对称性，也即狭义的洛伦兹变换成立。

现代物理学约定光速是一个常数 c ，有三层含义，即约定光速均匀、各向同性，而且与光源相对于观测者的运动无关。这就相当于要求时间和空间均匀、空间各向同性，而且光速不变原理成立。也就是说要求时空满足庞加莱对称性。

有引力场的时空不具有整体的庞加莱对称性。然而，测量是局域的，我们只需要约定局部光速，约定光速在每一时空点的邻域均匀各向同性，而且是一个常数 c 。这一约定对应局域庞加莱不变性。

广义相对论中的测量一定是局域的。我们“对钟”所用的小步雷达法，建立在局域测量的基础上。约定“局域光速”，相当于约定弯曲时空中存在局域庞加莱对称性。这与引力规范理论的主张是一致的。在引力规范理论中，引力场可以看作时空庞加莱对称性局域化而产生的补偿场——规范场。我们这里关于时间测量的讨论，似乎支持广义相对论是一个以庞加莱群为基础的引力规范理论。

从上述讨论可以看出下列关系：

对“真空中光速”的约定 \Leftrightarrow 对“时空对称性”的约定

约定真空中光速是常数 $c \Leftrightarrow$ 约定时空具有庞加莱对称性

光速的均匀性 \Leftrightarrow 时间、空间平移不变性 \Leftrightarrow 能量守恒、动量守恒

光速的各向同性 \Leftrightarrow 空间转动不变性 \Leftrightarrow 角动量守恒

光速不变原理 \Leftrightarrow 时空 Boost 不变性 \Leftrightarrow 狭义洛伦兹变换

对“好钟”的约定，可以看作对“时空对称性”（特别是时间对称性）的约定。均匀流逝的时间，可以保证能量守恒。各种时空对称性，可以保证物理规律简单。而对“时空对称性”的约定，相当于对光速的约定。

因此，我们只采用“约定光速”的办法，就可以定义异地时钟的同步和相继时间段的相等。“好钟”的存在，不再是一个约定，而只是对光速约定的一个推论。



绘画：张其

第十章 千古难题：时间是什么

在本章中，作者将把哲学家和物理学家们关于时间的杰出论述介绍给读者。

难以回答的问题

时间是什么？这个看似人人皆知的问题，回答起来却万分困难。

公元4世纪，早期基督教思想家圣·奥古斯丁在他的《忏悔录》中写下了一句至理名言：“时间是什么，人不问我，我很清楚，一旦问起，我便茫然。”

现代自然科学认为，时间和空间都是物质延展性的表现。时间是一维的，空间是三维的。静电场强与距离的平方成反比，支持了空间的三维性。也就是说，空间的三维性得到库仑定律的支持。时间的一维性则缺乏有力的支撑，只不过是人们由经验得到的结论，人们似乎很难想象时间会高于一维。

时间是一维的绵延；空间是三维的广延。绵延与广延都属于延展。然而绵延与广延不同，它还有“流逝”的含义。绵延的这种“流逝性”，在物理学上表现为自然过程的不可逆性，表现为热力学第二定律。流逝性的存在使得时间概念比空间概念更为复杂，因而也引起了哲学家和科学家的更多注意。

千百年来，不少伟大的思想家对时间发表了许多精彩的见解，进行了许多深奥的争论。这些见解和争论，加深了人类对时间的认识，给我们带来了光明，也带来了更多的疑义。

时间是永恒的映像

人类的时间概念来源于对事件先后顺序的排列，对因果关系的了解，对昼夜交替和季节变迁的认识。当人类文明发展到一定程度的时候，上述感性认识逐渐升华为理论。

柏拉图认为真实的“实在世界”是“理念”，我们感受和接触到的万物和宇宙都不过是“理念”的“影子”。理念完美而永恒，它不存在于宇宙和时空中。万物和宇宙则是不完美的，处于不断变化中。柏拉图认为造物主给“永恒”创造了一个“动态的相似物”，用以描述变化的宇宙和万物。那就是“时间”。他认为，时间是“永恒”的映像，是“永恒”的动态相似物；时间不停地流逝，模仿着“永恒”。

时间是运动的计数

柏拉图最优秀的学生亚里士多德，把老师的理论倒了过来，认为真实存在的不是“理念”，而是万物和宇宙组成的客观世界。对此柏拉图感到十分痛苦，为此亚里士多德说了一句千古流芳的佳话：“吾爱吾师，吾尤爱真理。”

在把老师的哲学观点倒过来的同时，亚里士多德也对时间发表了不同的见解。他认为“时间是运动的计数”，是“运动和运动持续量的量度”。时间概念的出现，使运动的测量成为可能，使我们可以区分快慢和静止。

无始无终的循环的时间

柏拉图认为时间无始无终，循环流逝，36000年为一个周期。亚里士多德也认为时间循环流逝，周而复始。

事实上，古希腊学术界始终认为时间是循环的，从已知最早的哲学鼻祖泰勒斯开始，到毕达哥拉斯、柏拉图、亚里士多德，他们都坚持或强或弱的循环时间观，并把时间与周期运动联系起来。

毕达哥拉斯学派说时间是天球，柏拉图说时间是天球的运动。亚里士多德认为时间是运动的计数，也把时间与天球的运动挂钩。

古印度同样是循环的时间观点占主导地位，这一观点也体现在佛教的教义中，对中国有一定的影响。

螺旋形发展的时间

然而，影响中国古代哲学思想的不仅有释家（佛教），而且有儒家（孔孟学说）和道家（老庄哲学）。

“子在川上曰：逝者如斯夫，不舍昼夜。”

“逝者”就是时间。孔夫子把时间比作永恒流逝的河流。他强调流逝，就是强调时间的不可逆性，这是他的高明之处。

我们从一些古诗词中也可以看出一些文人的时间观。中国文人的头脑中往往混杂着释家、儒家和道家的思想，他们认为时间不是简单的循环、重复，而是螺旋形发展的。首先是发展、流逝，其次又有一定程度的循环往复。例如唐朝诗人刘庭之（即刘希夷）的诗句

年年岁岁花相似，

岁岁年年人不同。

再如宋代文人晏殊的词

一曲新词酒一杯，

去年天气旧亭台，（循环的相似）

夕阳西下几时回。（时间不停流逝）

无可奈何花落去，（万物与时俱进）

似曾相识燕归来，（循环的相似）

小园香径独徘徊。

有始有终的线性时间

基督教的诞生对人类的时间观产生了重大影响。《圣经》认为世界万物和人类都是上帝创造的。上帝不在时间中，上帝在创造世界的同时创造了时间，因而时间有一个开端，还很可能有一个结束（世界末日）。时间不是循环的，而是线性演化的。

神学家们还明确指出了世界开始的时间，那里应该是时间的原点。宗教改革的创始人马丁·路德认为，上帝创造世界是在公元前4000年。

然而，后来开普勒根据天文研究发现，原来所认为的耶稣诞生之年有误，圣诞年差了4年，耶稣实际诞生于公元前4年。

圣经编年史权威、爱尔兰大主教乌塞尔根据路德和开普勒的研究，最终确定上帝创造世界是在公元前4004年10月22日下午8点。这一刻就是时间开始的原点。

牛顿与莱布尼茨的争论

文艺复兴和哥白尼的日心说划破了欧洲中世纪封建社会的黑暗，使古希腊的科学与艺术重见天日。人们重新开始了对哲学和科学的探讨。

牛顿的老师巴罗对时空有许多精辟的见解，牛顿把巴罗的思想加以发展提升，形成完整的绝对时空观。他认为存在不依赖于物质和运动的绝对时间和绝对空间。

牛顿写道：“绝对的、真实的和数学的时间，按其固有的特性均匀地流逝，与一切外在事物无关，又名绵延；相对的、表观的和通常的时间，是可悉知和外在的对运动之延续之度量，它常常用来代替真实的时间，如一小时，一天，一个月，一年。”

牛顿曾提出水桶实验来论证绝对空间的存在，但从来没有提出过任何实验来论证绝对时间的存在。他认为我们通常谈论、测量的时间都不是真实的绝对时间，而只是绝对时间的一种代替物——相对时间（表观时间）。相对时间只不过是“运动延续的度量”。

因此牛顿主张用运动来度量时间，但他认识到，“可能并不存在一种运动可以用来准确地测量时间”，“所有的运动可能都是加速的或减速的，但绝对时间的流逝却不会有所改变”。

在牛顿看来，绝对时间是一条无头无尾、始终如一的河流，没有“源头”，也没有涨落和波涛（这一点与孔夫子的想法非常接近）。时间除了均匀流逝的属性之外，没有其他属性。

也就是说，牛顿认为绝对时间是均匀的，有方向的，没有起点和终

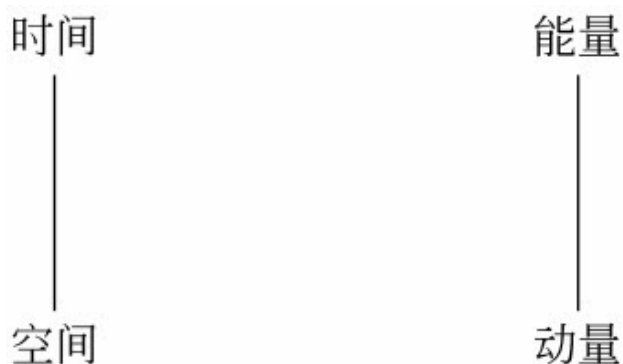
点的，永远存在的“河流”。如果物质消失了，时间和空间还会继续存在。

和牛顿同为微积分创建者，在发明权上与牛顿争吵不休的莱布尼茨，在对时空的看法上也与牛顿针锋相对。

莱布尼茨认为根本就不存在什么绝对空间和绝对时间，时间和空间都是相对的。空间是物体和现象有序性的表现方式，时间是相继发生的现象的罗列。时间和空间都不能脱离物质客体而独立存在。物质消失了，时间和空间也就消失了。

相对论的时空观

爱因斯坦的相对论，掀开了人类认识时空的新篇章。爱因斯坦认为时间是相对的，空间也是相对的，但时空作为一个整体是绝对的。能量是相对的，动量也是相对的，但能量-动量作为一个整体是绝对的。他的狭义相对论，建立起了时间与空间的联系，能量与动量的联系：



在广义相对论中，爱因斯坦进一步建立起时空与物质之间的联系



他认为物质的存在会造成时空的弯曲，弯曲的时空又会反作用于其中的物质，影响物质的运动。在广义相对论中，作为“演员”的物质和作为“舞台”的时空，不再是互不相关的，而是相互影响的。

不过，在相对论中，物质消失后，时空不会消失，时空依然存在，

只不过由弯曲变成了平直。

爱因斯坦晚年对时空的看法

晚年的爱因斯坦，曾经表达过对上述图像的不满，他写道：

“时间-空间未必能看成是可以脱离物质世界的真实客体而独立存在的东西。并不是物体存在于空间中，而是这些物体具有空间广延性。这样看来，‘关于一无所有的空间’的概念，就失去了意义。”

我们看到，晚年的爱因斯坦对时空有了新的看法。他认为：时空是物质伸张性和广延性的表现，不存在一无所有的时空，没有物质就没有时空，时空与物质同存同灭。

爱因斯坦的哲学观点，走在了他的物理理论（狭义相对论与广义相对论）的前面。从上述哲学观点中，不难看出莱布尼茨等人对他的影响。值得注意的是，目前建立量子引力理论的方案中，已有一部分学者采纳了爱因斯坦最后的哲学观点。

时空的泡沫与浪花

相对论极大地改变了人类的时空观念，量子论也在影响人类对时空的认识。宏观上看来平坦均匀的时空，在微观看来却存在量子涨落。不确定关系告诉我们，考察的时空区域越小，感到的量子涨落越激烈。

这就像飞机上的人观察大海一样。当飞机飞得很高时，飞行员觉得海面平静而光滑，降低飞行高度后，飞行员注意到海面上存在微波与涟漪。当飞机贴近海面时，飞行员会看到泡沫与浪花。

因此，相对论界普遍认为，在宇宙诞生的极早期，由于时空区域极小，可能存在着猛烈的时空量子涨落。在那段时期，时空的泡沫与浪花有可能形成复杂的拓扑结构，例如虫洞和多连通的宇宙。

时空的这类复杂结构，有可能伴随着宇宙的膨胀和演化，一直保留到今天。而且，量子引力理论告诉我们，在微观尺度下，时空有可能高于4维，很可能是10维或者11维。

孔夫子和牛顿描述过均匀流逝的时间之河。量子论告诉我们，近距离观察时，河面并不平静，存在着泡沫和浪花（图10-1）。

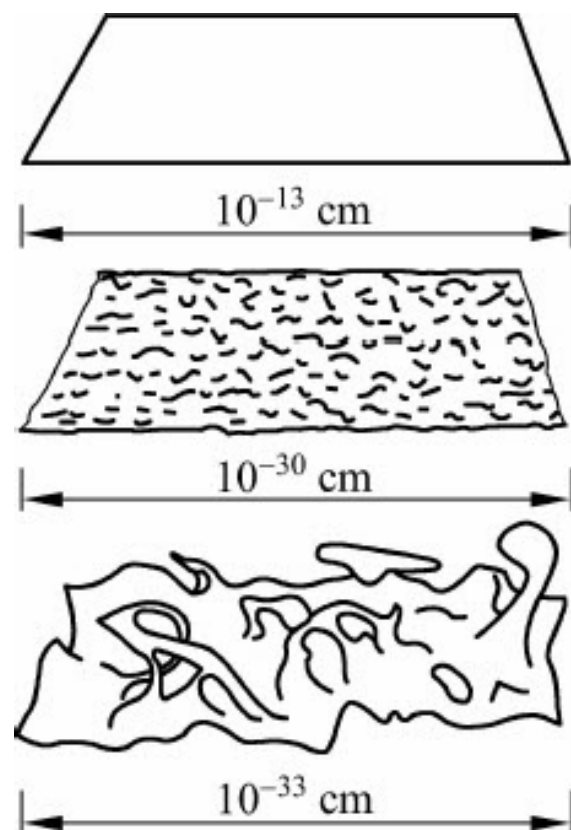


图10-1 时空涨落的泡沫与浪花

是否存在“虚时间”？

量子力学诞生的初期，人们就研究过势垒贯穿。一般认为势垒贯穿是一个瞬时过程，不需要时间。

最近我们在对黑洞隧道效应的研究中，发现弯曲时空中的势垒贯穿也是一种瞬时过程。在虫洞和时空隧道的研究中，人们也谈到穿越欧几里得虫洞是一个瞬时过程。

这类瞬时过程的物理本质是什么？是否意味着虚时间过程的存在？这是目前还不十分清楚的问题。

物理学把时间“空间化”了吗？

我们看到，物理学的发展大大推动了时空观念的发展，促进了人类对时间和空间的认识。可以说，物理学对哲学功不可没。

然而，近年来，物理学中的时间观念却受到哲学家的猛烈批判。最尖锐的指责是“物理学家把时间概念空间化”了，就是只强调了时间与空间的共性，没有强调两者的区别。主要是没有强调时间的“流逝性”，即“不可逆性”。

这种批评意见对物理学的大多数领域是有道理的。包括各种场论、相对论和量子论在内的几乎物理学的所有领域，都只考虑了时间的均匀性、持续性，却回避了时间的流逝性。

然而，在热力学领域，这一批评意见不成立。热力学第二定律明确指出了时间的方向性、流逝性和不可逆性。正是这条定律，成为了哲学界探讨时间流逝性的科学基础，而且到目前为止，它几乎是唯一的科学基础。

时空弯曲与时间之矢

物理学有两个分支非常值得注意，一个是广义相对论，另一个是热力学。

除广义相对论以外所有的物理领域（包括热力学），都把时空看作平直的背景，不考虑物质运动与时空背景之间的相互作用，只有广义相对论认为物质与时空背景有相互作用，物质的存在会使时空背景弯曲，弯曲的时空背景又会反过来影响物质的运动。

除热力学之外的所有物理领域（包括相对论），都不考虑时间的流逝性，都认为物理理论在时间反演（或CPT变换）下不变。只有热力学第二定律显示出时间箭头，指出了时间的流逝性和不可逆性。

看来，未来的物理学家应该把“时空弯曲”和“时间箭头”的考虑渗透到本学科的所有领域，形成一个能够同时反映时间均匀性、持续性和流逝性的统一的物理理论。

熵：时间流逝的计量

时间有两个基本特性：流逝性和测度性。空间只有测度性没有流逝性。因此，时间与空间最大的区别，就是它的流逝性，一去不复返性。

作为物理学最重要定律之一的热力学第二定律，就是要指明时间的流逝性。这条定律所强调的核心——自然过程的不可逆性，清楚地指明了时间流逝的方向。这条定律确定了一个物理量——“熵”，它是混乱度的量度，可以用它来定量地刻画时间的流逝。

自然过程的不可逆性显示了时间箭头的存在，这种时间之箭指向熵增加的方向，无论是摩擦生热等自然过程还是量子力学中的“测量”过程，都指向熵增加的方向，即指向混乱度增加的方向。

时间意味着创造

除去上述明显的从有序到无序转化的过程外，人们还注意到另一类好像完全相反的过程，如生物进化和社会进化。这类过程从无序向有序、从简单向复杂发展，似乎给出了相反的时间箭头。柏格森等哲学家注意到上述过程，因而强调时间意味着创造，意味着进化。

柏格森毕业于巴黎高等师范学院，毕业后在著名科学家帕斯卡的故乡克莱蒙菲中学当教师，后来又做过大学的临时讲师。

“一个暮春时节的黄昏，25岁的教师柏格森散步走到克莱蒙菲城郊。这是法兰西腹地的高原地带，漫山遍野生长着高大的树木，西天的晚霞在万里长空中向东天铺洒开来，远处卢瓦尔河的支流在奔流不息。柏格森站在高处，目睹着河水奔流、树木摇曳、晚霞飘逝，突然对时光之逝产生了一个非常震惊的感觉。”（引自吴国盛《时间的观念》）

这一震撼促使柏格森在当教师的同时，展开了对“时间性质”的研究，完成了他的第一部专著《时间与自由意志》。他提出了以“绵延”为核心的时间理论。

柏格森认为“时间与人的意识和直觉有关”的观点虽然值得商榷，但他强调时间的“动”，时间的“流逝”，把时间看作创造，看作进化的思想却是难能可贵的。柏格森用激情荡漾、技巧高超的演讲，把他的观点广泛传播给大众，引发了震撼的效果，许多年青人深受他的影响。

柏格森对普里高津的启示

著名的物理学家普里高津后来回忆道：“在我年轻的时候，我就读了许多哲学著作，在阅读柏格森的《创造进化论》时所感受到的魔力至今记忆犹新。尤其是他评论的这样一句话：‘我们越是深入地分析时间的自然性质，我们就会越加懂得时间的延续就意味着发明，意味着新形式的创造，意味着一切新鲜事物连续不断地产生。’这句话对我来说似乎包含着一个虽然还难以确定，但是却具有重要作用的启示。”

柏格森把时间看作“创造”与“进化”的思想，是跟我们看到的“生物进化”和“社会进化”相一致的。在生物与社会的进化中，我们看到的是从无序到有序，从简单到复杂的演进。

这些现象似乎与热力学第二定律显示的演化方向相反。第二定律虽然告诉我们时间是“动”的，是“流逝”的，但这种“动”和“流逝”的过程是熵增加的过程，混乱度增加的过程，系统从有序向无序、从复杂向简单演变的过程。

正是柏格森观点与第二定律表面上的矛盾，给了普里高津以重大启示。他在继承物理学衣钵的同时，也接受了柏格森的极具启发性的哲学思想。从此之后，普里高津把自己的毕生精力奉献给了对“不可逆性”和“时间之矢”的研究。

耗散结构

普里高津注意到热力学第二定律推出的熵增加原理，针对的是孤立系统或绝热系统（绝热系统是指该系统与外界不发生热交换），这样的系统与外界没有熵的交换。而生物体、社会单元都是开放系统，与外界有熵交换。

普里高津提出“耗散结构”的概念（这一概念，现在被称为非平衡系统的自组织结构），它指的是一种远离平衡态的开放系统自发地形成有组织的有序结构的情况。

这种系统不断地从外界吸入低熵物质，排出高熵物质，总的效果相当于从外界输入“负熵”。这种系统内部经历着不可逆过程，不断有熵产生，而“负熵”的输入抵消了新产生的熵，从而使系统的熵维持大致不变，因而系统本身相对稳定。

例如一个动物，内部的新陈代谢是不可逆过程，不断产生熵。但该动物不断吃进食物，排出粪便和尿液等。食物中的分子排列比较有序，是低熵物质；排出物（粪便等）中分子排列比较无序，混乱度大，是高熵物质。吃进低熵的食物，排出高熵的排出物，就相当于吸进了“负熵”。这些负熵可以抵消动物体内部新陈代谢增加的熵，从而维持动物体的熵大体不变，因而保持了动物体内的相对稳定。动物体就是一种耗散结构，它是一个不断从外界吸入负熵的开放系统。

一座城市，在生存过程中不断产生垃圾。必须不断向城市运进各种食物和用品（低熵的物质），不断从城中运出垃圾、排出污水（高熵物质），城市才得以生存。所以说，城市也是一种耗散结构，是一个不断

从外界吸入负熵的开放系统。

耗散结构理论，为解释生物和社会的生存与进化打下了基础。我们看到，这种不断进化的客体，从外界吸收负熵，但“客体”本身与“外界”的总熵仍是增加的，因此并不与热力学第二定律矛盾。从普里高津的耗散结构理论可以看出，柏格森把时间看作“创造”与“进化”的观点，并不与热力学第二定律抵触。

普里高津从热力学第二定律，从物理学给出了时间“意味着创造”，“意味着进化”的结论。在《从存在到演化》一书中普里高津指出：“我们看到了某些最近的结论与柏格森、怀特海和海德格尔等哲学家的预期有多么接近。主要的区别是，在他们看来，这样的结论可能只是由于与科学的冲突而得到的，而我们现在把这些结论看作可以说是从科学研究的内部得出的。”

宇宙学的时间箭头

物理学和天文学还给出了另一种时间箭头——宇宙学箭头。许多文章介绍了大爆炸宇宙学。这一理论指出，宇宙从奇点开始，大爆炸后不断膨胀，不断降温。虽然对于宇宙发展的不同阶段，膨胀速度如何，膨胀是否经历过减速和加速，大家意见尚未统一，但宇宙从开始到现在，一直在膨胀，一直在降温，却是公认的结论。大家都认为，随着宇宙膨胀，宇宙的熵在增加，与热力学第二定律一致。

无限无边的宇宙会一直膨胀下去，一直降温，熵也会一直增长下去。有限无边的宇宙膨胀到一个最大体积后会收缩，在收缩阶段，温度会升高，至于宇宙中的熵如何变化，有两种意见，一种认为熵会减少，甚至有人认为这意味着时间会倒流。这时热力学第二定律会被破坏，或者第二定律这时可能改为“熵减少”的说法。不过绝大多数人认为，在宇宙收缩阶段，熵仍会增加，第二定律仍然成立，时间箭头绝不会倒过来。

近年来彭罗斯等人从微分几何和广义相对论角度做了有意思的探讨，他们试图用一个叫做“魏尔张量”的几何量来刻画熵，论证在宇宙的收缩阶段，熵将继续增加，大塌缩的终极奇点不同于大爆炸的初始奇点，二者并不对称，终极奇点的熵高于初始奇点的熵。

心理学的时间箭头

还有一个时间箭头是心理学时间箭头。其含义是：人类只能回忆过去，不能忆及未来。这表明，从心理的角度看，“过去”与“未来”是不对称的。实际上我们对未来会如何，可以设想多种可能性，但只能实现其中一种，而且不可逆转，也不能反悔。时间的进展，意味着不断地把不确定的“未来”，转化成为确定的“现在”，继而成为确定的“过去”。

这种情况有点像量子力学中的测量，测量使量子的纯态转化为混合态，从多个纯态的叠加塌缩到一个确定的态。在量子力学测量过程中，熵在增加。从过去向未来的发展，也肯定伴随着熵增加。由此看来，心理学的时间箭头只不过是热力学时间箭头的反映，是热力学第二定律的另一种表现。

从上面的讨论可以看出，热力学第二定律所给出的“不可逆性”是所有各种时间箭头的科学基础。因此，不仅生物学、社会学的时间箭头，而且宇宙学、甚至心理学的时间箭头，都不过是热力学时间箭头的推论。所以我们可以说，物理学没有把时间空间化。而且，能够指出时间流逝性的科学，唯有物理学。

然而，物理界也应从哲学界的批评中吸取正确的成分，注意努力把时间的流逝性渗透到自己的各个分支中去。

时间之矢何以产生？

另一个问题是：以分子运动论为代表的统计力学描述的微观粒子运动都是可逆的力学运动，何以大量粒子的行为会呈现出宏观的不可逆性？呈现出时间箭头？从玻耳兹曼开始就对这一问题进行过探讨。玻耳兹曼给出了熵的表达式，给出了研究碰撞的玻耳兹曼方程和H定理，使这一问题的研究取得了重要进展。然而问题还没有解决。

普里高津等人推动的“对时间的再发现”的研究，在耗散结构、混沌等领域都取得了进展，但对时间之箭何以出现，仍然不能给出解释。

值得注意的一件工作是普里高津曾经耗费大量心血的、一次没有成功的尝试。他试图用非么正变换来引进不可逆性，使微观的可逆过程过渡到宏观的不可逆过程。为此他与密斯纳等做了大量的数学、物理研究，并取得了一些进展。然而在他们的工作中，需要在两个对称的演化序列中做一次初始的选择，有人怀疑正是这一选择暗含了“不可逆性”，因此不能认为他们的工作已经成功。不过，笔者认为他们采用非么正变换来表述不可逆性的做法值得借鉴。

在讨论黑洞信息疑难的时候，人们注意到信息丢失（熵增加）会破坏量子理论的么正性。由于信息可以看做负熵，信息的缺失意味着熵增加，可见，把“不可逆性”与非么正变换相联系是一个正确的方向。

从原则上讲，不可逆性并不需要有一个根源，它本身就是根源。热力学第二定律本身是一条基本公理，根本不需要导出，我们需要解决的问题只是：大量微观粒子的可逆运动何以在宏观上呈现出不可逆性，呈现出时间箭头。

展望

对广义相对论和黑洞的研究，引导我们走向了对时间之河的探索。我们在研究中发现：时间与热力学之间似乎存在着比迄今所知更为深刻的内在联系。

在我们看来，热力学的四条定律都与时间的性质有关：第零定律表明，时间是可以定义的；第一定律表明，时间是均匀的；第二定律表明，时间是流逝的，有方向的；第三定律表明，时间是无穷无尽的，既没有开始，又没有结束。

时间就像一条既没有源头，又没有终点的河流，向着特定的方向均匀地流逝着，远看起来，它十分平静，近看起来，会发现它有量子效应导致的涟漪、波涛和浪花。

人类对时间的认识在不断深化，但是至今仍然无法回答“时间究竟是什么”这一难题。

主要参考书目

科学著作与教材

[1] 爱因斯坦A. 相对论的意义 [M] . 李灏, 译. 北京: 科学出版社, 1961.

[2] Newton I. Mathematical Principles of Natural Philosophy [M] . Cambridge: Cambridge University Press, 1934.

[3] 朗道, 栗弗席兹. 场论 [M] .8版.鲁欣, 任朗, 袁炳南, 译. 邹振隆, 校. 北京: 高等教育出版社, 2012.

[4] 温伯格S. 引力论和宇宙论 [M] .邹振隆, 张历宁, 等, 译. 北京: 科学出版社, 1980.

[5] 刘辽, 赵峥. 广义相对论 [M] . 2版. 北京: 高等教育出版社, 2004.

[6] 梁灿彬, 周彬. 微分几何入门与广义相对论 [M] . 2版.北京: 科学出版社, 2006.

[7] 俞允强. 广义相对论引论 [M] . 北京: 北京大学出版社, 1987.

[8] 须重明, 吴雪君. 广义相对论与现代宇宙学 [M] . 南京: 南京师范大学出版社, 1999.

[9] 王允然, 鲁菲尼R.相对论天体物理的基本概念 [M] . 上海: 上海科学技术出版社, 1981.

[10] 瓦尼安H C, 鲁菲尼R. 引力与时空 [M] . 向守平, 冯珑珑, 译. 北京: 科学出版社, 2006.

[11] 赵峥. 黑洞的热性质与时空奇异性 [M] . 北京: 北京师范大学出版社, 1999.

[12] 刘辽, 赵峥, 田贵花, 等. 黑洞与时间的性质 [M] . 北京: 北京大学出版社, 2008.

[13] 赵峥. 黑洞与弯曲的时空 [M] . 太原: 山西科学技术出版社, 2000.

[14] 赵峥, 刘文彪. 广义相对论基础 [M] . 北京: 清华大学出版社, 2010.

[15] Wald R M. General Relativity [M] . Chicago: The University of Chicago Press, 1984.

[16] Hawking S W, Ellis G F R. The Large Scale Structure of Space-time [M] . Cambridge: Cambridge University Press, 1973.

[17] Birrell N D, Davies P C W. Quantum Fields in Curved Space [M] . Cambridge: Cambridge University Press, 1982.

[18] Rindler W. Essential Relativity [M] . New York: Springer-Verlag, 1977.

[19] Misner C W, Thorne K S, Wheeler J A. Gravitation [M] . San Francisco: Freeman W H Company, 1973.

[20] 王永久. 经典黑洞与量子黑洞 [M]. 北京: 科学出版社, 2008.

[21] Prigogine I. From Being to Becoming [M]. San Francisco: Freeman W H and Company, 1980.

[22] 普里高津. 从存在到演化 [M]. 曾庆宏, 严士健, 马本堃, 等, 译. 上海: 上海科学技术出版社, 1986.

[23] 胡中为, 萧耐园, 朱慈盛. 天文学教程 [M]. 北京: 高等教育出版社, 2003.

科普类

[24] 爱因斯坦A. 狭义与广义相对论浅说 [M]. 杨润殷, 译. 上海: 上海科学技术出版社, 1964.

[25] 霍金S W. 时间简史 [M]. 许明贤, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1994.

[26] 彭罗斯R. 皇帝新脑 [M]. 许明贤, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1994.

[27] 霍金S W. 霍金讲演录 [M]. 杜欣欣, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1994.

[28] 霍金S, 彭罗斯R. 时空本性 [M]. 杜欣欣, 吴忠超, 译. 长沙: 湖南科学技术出版社, 1996.

[29] 陶宏. 每月之星 [M]. 上海: 开明书店, 1949.

[30] 王允然, 褚耀泉. 从牛顿定律到爱因斯坦相对论 [M]. 北京: 科学出版社, 1982.

[31] 卢米涅J P. 黑洞 [M]. 卢炬甫, 译. 长沙: 湖南科学技术出版社, 1997.

[32] 纳里卡J V. 轻松话引力 [M]. 卢炬甫, 译. 长沙: 湖南教育出版社, 2000.

[33] 徐一鸿. 爱因斯坦的玩具: 探寻宇宙和引力的秘密 [M]. 张礼, 译. 北京: 清华大学出版社, 2013.

[34] 卢昌海. 从奇点到虫洞: 广义相对论专题选讲 [M]. 北京: 清华大学出版社, 2013.

[35] 索恩K S. 黑洞与时间弯曲 [M]. 李泳, 译. 长沙: 湖南科学技术出版社, 2000.

[36] 吴国盛. 时间的观念 [M]. 北京: 中国社会科学出版社, 1996.

[37] 诺维科夫. 时间之河 [M]. 吴王杰, 陆雪莹, 译. 上海: 上海科学技术出版社, 2001.

[38] 保罗·戴维斯. 关于时间 [M]. 崔存明, 译. 长春: 吉林人民出版社, 2002.

[39] 彭加勒. 科学与假设 [M]. 李醒民, 译. 北京: 商务印书馆, 2006.

[40] 彭加勒. 科学的价值 [M]. 李醒民, 译. 北京: 商务印书馆, 2007.

[41] 赵峥. 探求上帝的秘密 [M]. 北京: 北京师范大学出版社, 2009.

[42] 赵峥. 物理学与人类文明十六讲 [M]. 北京: 高等教育出版社, 2008.

[43] 邓乃平. 懂一点相对论 [M]. 北京: 中国青年出版社, 1979.

[44] 赵峥. 相对论百问 [M]. 北京: 北京师范大学出版社, 2010.

[45] 赵峥. 物含妙理总堪寻 [M]. 北京: 清华大学出版社, 2013.

[46] 郭中一. 科学, 从好奇开始 [M]. 台北: 文经社, 2005.

科学史

[47] 郭奕玲, 沈慧君. 物理学史 [M]. 2版. 北京: 清华大学出版社, 2005.

[48] Pais A. The science and the life of Albert Einstein [M]. Oxford: Oxford Univ. Press, 1982.

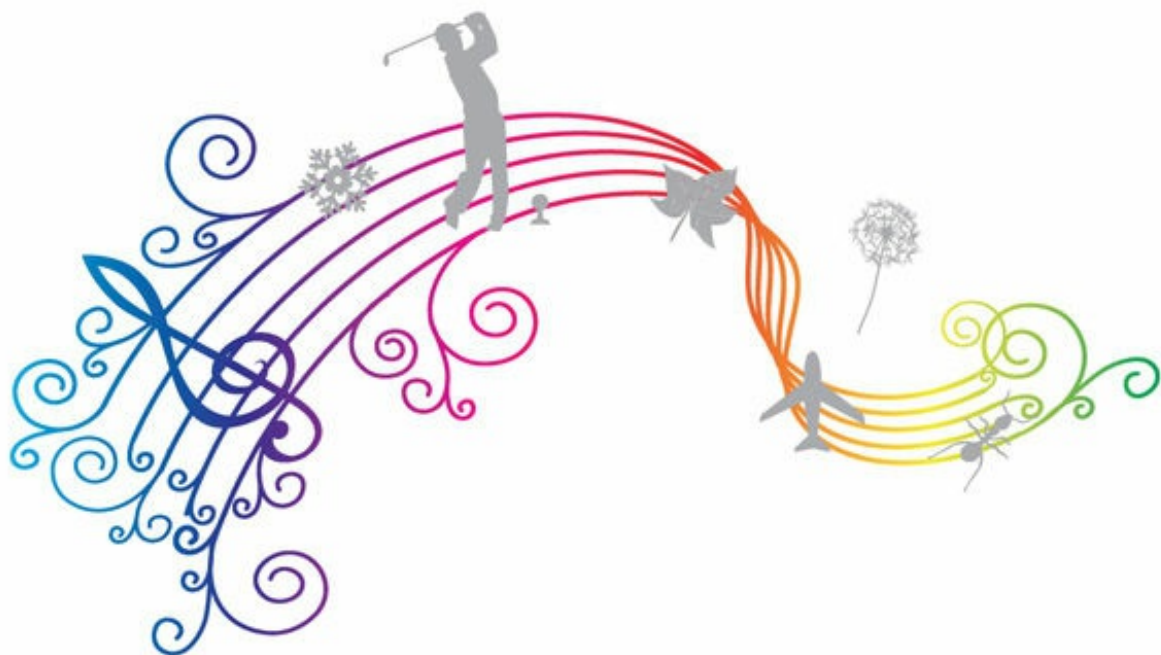
[49] 派斯. 爱因斯坦传 [M]. 方在庆, 李勇, 等, 译. 北京: 商务印书馆, 2006.

[50] 乔治·伽莫夫. 物理学发展史 [M]. 高士圻, 译. 北京: 商务印书馆, 1981.

[51] 沃尔特·艾萨克森. 爱因斯坦传 [M]. 张卜天, 译. 长沙: 湖南科学技术出版社, 2012.

[52] 怀特M, 格里宾J. 斯蒂芬·霍金的科学生涯 [M]. 洪伟, 译. 周环, 校. 上海: 上海译文出版社, 1997.

理解科学丛书
大众科普著作



吴子牛 编著

运动的旋律与 变化的世界

清华大学出版社

运动的旋律与变化的世界

吴子牛 编著

清华大学出版社
北京

作者简介

吴子牛 北京航空学院学士毕业，法国巴黎第六大学博士毕业，现为清华大学教授和教育部长江学者。担任过多个方面的专家组成员和多家企业的顾问，在多个专业领域有一定的建树，在国际期刊上发表了约百篇与数学、工程和自然相关的论文，包括火旋风以及昆虫行为的论文。将文字、旋涡、流行病、雾化和人口等的演化当作统一规律的科学论文被国际期刊《熵》邀请发表。在传统文化、现代教育和应用软件方面有广泛的兴趣。

用徘徊于通俗描述和专业逻辑之间的折中语言包括适当的拟人化语言，给读者展示和解释一些诸如雨滴与蚊子、钟爱数字7、飞机与昆虫、时间、蒲公英恋南风、高尔夫球、棒棒糖的融化、雪花与树叶飞舞等数百个常见现象的简单、简约和简美的一面，将某些看似复杂不可理解的现象用简单现象叠加或类比，深奥的原理用省时、省力和省料那样的直觉性语言表述，并适当关联出一些与文字、体育、生物、气象、音乐、心理学、数学和科技等方面的有用常识。

本书分“动的世界”“柔情似水”和“空中旋律”三大篇，是一部集知识性、趣味性和科学性于一体的综合科普读物。适合大中学生和具有各种知识结构的读者群，尤其适合那些喜欢用非专业思维、语言和大视野去理解科学、自然和其他现象的读者。也可用于有一定阅读和讲解能力的读者给有好奇心的少儿讲解知识。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

运动的旋律与变化的世界 / 吴子牛编著. —北京：清华大学出版社，
2017

ISBN 978-7-302-46992-6

I. ①运... II. ①吴... III. ①科学知识—普及读物 IV. ①Z228

中国版本图书馆CIP数据核字（2017）第100135号

责任编辑：佟丽霞

封面设计：常雪影

责任校对：赵丽敏

责任印制：杨 艳

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印 刷 者：北京富博印刷有限公司

装 订 者：北京市密云县京文制本装订厂

经 销：全国新华书店

开 本：145mm×210mm

印 张：11.25

字 数：280千字

版 次：2017年7月第1版

印 次：2017年7月第1次印刷

印 数：1-3000

产品编号：073438-01

序 言

《运动的旋律与变化的世界》分“动的世界”“柔情似水”和“空中旋律”三大篇。每一篇标题表面上看显得很专，实际上牵引出很多其他问题，如数字7、时间、几何图形、圆周率、数学、少儿娃娃浪、大气污染、音乐、说话的技巧、针尖效应等。用大主题关联不同现象，用大视野看问题，正是本书的目的之一。例如，针尖效应出现在最后的雪花问题之中，但针尖效应又对理解植物叶子为何有的长有的短这类现象有帮助。因此表面上看很专的具体问题，往往牵引出更具普适意义的现象。

从结构上，本书每一篇分为四大节，每一节有五个主题，每个主题涉及的具体问题也可能包含许多相似或相关的现象。虽然各个主题具有一定的独立性，但前后描述时所基于的知识有一定的关联性，这种展现方式如同拴在抖动中的弦线上的珠子在一起舞动一样。有的强调趣味性和知识性，有的也包含很深的逻辑推理。部分逻辑推理虽然以启发式为主，但隐含的思维深度不一定亚于专业逻辑。这些有深度的内容也是为了适应那些喜欢深入思考的读者。

脍炙人口的科普读物将那些出现在厚厚的专业书中的知识提炼成人人都可以理解的内容，以现代小型读本形式出现。正是这些小型读本，使知识得到普及、求知欲得到激发，并给不同知识结构甚至不同行业的人群之间提供了认知世界的桥梁。

有关单一领域问题的科普读物不胜枚举，但更广义的科普读物并不

常见。许多表面上完全不同的问题，实际上具有共性。例如，打水漂时，水漂的弹跳与冲浪运动中的腾空几乎基于同样的道理，汉代发明的橹产生推进力的原理与现代飞机机翼产生升力的原理是一样的。又如，汉字笔画超过7笔后，具有相同笔画数目的汉字数目的增加速率就下降了，11笔左右汉字数目最大。这个11除以7得到的比值，其实不局限于汉字笔画数目的增长规律，许多其他现象都类似，比如说流行病爆发后，住院患者的数目就有类似规律。

许多现象，无论是科学、技术、经济还是文学艺术，至少从某种角度看，是朝着省时、省力、省料的方向发展演化，演化的结局往往是简单、简约、简美。例如，不管是什么类型的问题，对数目的选择一般是在少了不够和多了难以驾驭这两种之间博弈的结果，妥协的结果是数字“七”可能得到宠爱，如七言诗、七色光、七大洲、七种味觉、物理学有七个基本单位、七种生命形态等。

千变万化的世界中，许多问题、现象、知识和规律，看上去像散落的珠子。用一根弦把它们串起来，再抖动这根弦，我们就能看到它们相互关联与共性的一面。

《运动的旋律与变化的世界》正是一种这方面的尝试，通过适当采用直觉、类比和拟人化的手笔，给读者串联似地展示甚至解释数百个与我们的生活息息相关的问题与现象。即使不得不借用一些专业逻辑，也尽量将这种专业逻辑变为可以直觉关联的语言（站在专业角度看，当然适当牺牲了一些严谨性）。正是那些相互关联且道理相通的原则，将千变万化的运动和演化现象谱写成优美的旋律，使表面上复杂的大自然也有简单、简约和简美的一面。

奔跑中的腿和水面波浪中的水滴运动，显得多么复杂。如果看成同

时在做倒单摆运动和弹簧运动，就很容易理解了。一种称为水黽的小昆虫，借用腿上的小毛毛撑开水面，从而可以凌波微步和玩蹦床运动。蒲公英举着冠毛伞留在空中，因为有了又小又多的冠毛而能随风飘浮。都是小毛毛，前者憎水，后者恋风。鸟的翅膀横着飞就是产生升力的机翼，竖着拍就是产生推进的螺旋桨。以这些方式理解或关联知识和现象，正是本书的特点，期望让读者不需要掌握专业知识就能理解。

当然，不能纯粹为了好表述就来筛选那种显得又轻松又有趣的问题，也有一些显得很沉重和深奥的内容，如相对论效应、振动的世界、演化的普适规律、飞机气流和发动机。比如说，飞行的秘密专门有较为冗长的一节。站在科普一级，完全可以用有限的篇幅交代清楚，无非就是飞机和鸟有了机翼或翅膀，由迎角和弯度产生了上吸下举的差异性气流。然而，这部书并不是为了简单罗列知识和堆砌原理，而是希望借此引导读者用直觉进行非专业级的理解和思考，进而尽可能牵引出更多的常识来辅助理解和丰富知识。陈述历史可以将本书的篇幅翻倍，然而，陈述历史不是本书要突出的风格。对于每个问题，我们也不会采用统一的风格。比如说，开篇的打水漂，我们一开始就点出水漂弹跳的秘密。但有的问题，包含的原理贯穿于全文之中。

除了某些图片包含若干数学公式，本书正文主要只是文字描述和数据。提供大量的数据信息和可帮助理解的示意图，也是本书的一大特点。依据问题的性质，不同内容采用的语言风格可能不一样。有的适合用拟人化语言，有的则需要显得很书面很严肃。有的问题目前本来就不是特别清楚，反而适合用磕磕碰碰底气不足的语言。

大部分问题与我们的生活息息相关，这些问题包括澡盆涡、玩具陀螺、自行车、波浪、卫星、打水漂、墨西哥人浪、肥皂泡泡、棒棒糖、笔画、圆周率、飞机、高尔夫球、时间、高速列车、蚊子、潮汐、雪花

等。对这些知识的介绍与讨论，会关联出一些与文字、音乐、体育、生物、气象、数学和科技等方面的常识。

用小型读本涉及较为广泛的问题，正是为了体现现象具有碎片性而内涵具有关联性。因此，部分内容所涉猎的知识跨度、宽度或深度，会留有余韵或瑕疵，如此可为读者自己思考腾出或打开一些想象与思考空间。

本书是一部集知识性、趣味性和科学性于一体的综合科普读物。适合大中学生及以上具有各种知识结构的读者群。家长也可以将其中一些内容剥离出来用于给少儿讲故事。

吴子牛

2017年1月

目 录

[序 言](#)

[第一篇 动的世界](#)

[1.1 让形状和姿态疏通运动的路径](#)

- [1. 打水漂 魔幻迎水角](#)
- [2. 不修边幅的飞石](#)
- [3. 谁都钟爱流线型 远走高飞的高尔夫球](#)
- [4. 省力的橹与飞行的秘密 迎角的升力与抬头效应](#)
- [5. 带着气流旋转 弧线球](#)

[1.2 失稳、旋转与振动让世界更灿烂](#)

- [1. 失稳中的多姿多彩或险象重重](#)
- [2. 旋转助稳 令人痴迷的玩具陀螺](#)
- [3. 振动的世界 共振](#)
- [4. 极限飞盘 自行车](#)
- [5. 旋转的天体与优美的宇宙](#)

[1.3 运动与演化 过程的激荡与结局的完美](#)

- [1. 速度的消失带来损伤 蚊子会被雨滴砸坏吗](#)
- [2. 生长与衰减 动物大小 技术能力的演化](#)
- [3. 演化的普适规律 省时省力省料](#)
- [4. 演化规律能解释我们的世界吗](#)
- [5. 数字7](#)

[1.4 速度的限制 谁能把我们带往宇宙深处](#)

- [1. 移动了几倍体长或音速 尺度比例效应](#)
- [2. 太快也不安全 幸亏有音障、热障与光障](#)
- [3. 相对论效应 卫星定位 核能利用](#)

[4. 时间很慢也很快](#)

[5. 星际旅行 费米悖论](#)

[第二篇 柔情似水](#)

[2.1 柔情似水 凌波微步的小昆虫](#)

[1. 柔情似水](#)

[2. 绿叶上的水珠 纯水泡泡与肥皂泡泡](#)

[3. 鸡毛掸子一样的水黽腿 镊子一样的刚毛](#)

[4. 凌波微步 划桨运动 蹦床运动](#)

[5. 驱赶、威慑和求偶 通信与存储](#)

[2.2 风蚀地貌 圆与漩涡的生命力](#)

[1. 棒棒糖舔一千下 地貌的演化](#)

[2. 河道弯曲 牛轭湖 蜿蜒度](#)

[3. 圆周率 自然界中的常数 数学的魅力](#)

[4. 澡盆涡](#)

[5. 溪流、河道与海洋中的漩涡](#)

[2.3 轻抚水面生涟漪 强风破水摇巨浪](#)

[1. 蹲波队列 墨西哥人浪 幼儿园娃娃浪](#)

[2. 不揭短的涟漪用美丽衬托客人](#)

[3. 不插队不打架的涟漪波纹](#)

[4. 波速也会论短长 表面张力与重力的分工](#)

[5. 各种风速下的风浪](#)

[2.4 奇妙无比的水面世界](#)

[1. 无风不起（风）浪 无风三尺浪（涌）](#)

[2. 游泳池的震荡 湖震 反射与叠加](#)

[3. 海洋潮汐 钱塘江大潮 海啸](#)

[4. 物体与水波竞速 O形波与V字形波](#)

[5. 破浪而行 船与游泳形成的波浪](#)

[第三篇 空中旋律](#)

[3.1 绚丽多彩与刚柔并济的地球大气](#)

- [1. 大气的力量](#)
- [2. 神奇的自洁功能 环境污染](#)
- [3. 分子的运动 个体的行为与集体的表现](#)
- [4. 疏密有致 胀压有序 雷霆万钧](#)
- [5. 火旋风 空气中的旋涡 昆虫与鸟的旋涡](#)

[3.2 听懂世界](#)

- [1. 听不见与听得见的声音 蝴蝶与大象](#)
- [2. 烦人的噪声 高速物体的音爆](#)
- [3. 动听的声音 音乐与频率](#)
- [4. 声源的信息 口语与书面语言](#)
- [5. 风中的声音 移动物体的声音 追赶声音的感觉](#)

[3.3 飞行的奥秘](#)

- [1. 水遇窄道快 人（车）遇窄道慢](#)
- [2. 风中转掌知升力 优美的流线与和谐的声波](#)
- [3. 在机翼的气流中旅行](#)
- [4. 飞机与发动机](#)
- [5. 翼尖翻转涡 飞机起降频率 候鸟的迁徙](#)

[3.4 自然界的运动：一曲动感的旋律](#)

- [1. 鸟与昆虫的飞行秘密](#)
- [2. 蒲公英与降落伞 蝴蝶果与直升机](#)
- [3. 雨滴下落 流星自焚 航天飞机与返回舱再入](#)
- [4. 生物推进的相似性 鱼的游动](#)
- [5. 针尖效应 雪花 雪花铃](#)

[后记](#)

[返回总目录](#)

第一篇 动的世界

宇宙在膨胀、地球在自转、分子在运动、生物在演化，因此，我们处在一个动的世界之中。运动有形态，位置的更替、姿态的改变、速度的增减，谱写出动感的旋律。演化有节奏，数目的变动、大小的变化、范围的缩放，让大自然生生不息。

千变万化，却离不开自然规律的约束。简单、简约、简美，才是永恒的自然规律。物理学再复杂，也就七个基本物理量单位。简美的流线型物体，能顺利地避让挡路的空气和水，因此更能接近速度的极限。再复杂的演化现象，也会遵循省时、省力、省料甚至物极必反的原则。速度是一种能量，专业搏击手能击打出一千千克的打击力。直线运动如果无法及时释放能量，那么旋转就是不错的选择。雨滴休想消灭轻巧的蚊子，侧歪一下就躲掉了。

深奥的科学规律，当与我们的直觉能发生共鸣时，就成了我们理所当然的常识。于是，我们本能地会认为侧踹的足球会走弧线、带酒窝的高尔夫球飞得更远、旋转的陀螺能躲猫猫般地不倒、汉字的笔画不会太少也不会太多、速度高了会有限制。我们走进动的世界，探求隐喻在常识中的原因，会让我们的常识得到升华，会让我们发现更多的动感形态。

1.1 让形状和姿态疏通运动的路径

空中有小得看不见、多得数不清的空气分子，水中也是如此。空气和水也爱美。它们用事实告诉你，拥有顺从我的形状和姿态，就可以给你让道，甚至把你举起来，可以让你穿梭自如、风驰电掣。如果你不修边幅，长得像随便捡来的小石头一样，你将遭遇重重险阻。于是，高速列车、现代汽车和飞机乖乖地披上了流线型外衣。高尔夫球就不是流线型，只能用酒窝型凹槽把气流磕乱，磕得气流找不到方向，才能远走高飞。水与空气用同样的方式和穿梭的物体打交道，飞行的谜底居然从汉代发明的橹能找到答案。答案很简单：直接抬起你的头，或者带动空气旋转间接抬起你的头，看得高才能飞得远。抬头也不要过分，打水漂就有一个合适的迎水角。

1. 打水漂 魔幻迎水角

将一块扁平小石片贴水扔出去，让它在水面弹跳并激起一串串涟漪、一朵朵水花，这就是打水漂，作为民间活动历史悠久，作为国际赛事只有二十余载。边在空中飞行、边在“蜻蜓点水”，原理似乎深奥不可理解。它是怎么弹跳的？揭开其神秘面纱，既可掌握增加弹跳次数的技巧，也可启发水上和空中的一些应用。原来，在入水时，水漂在腹部制造一个冲浪，凭惯性做冲浪腾空，当然就能弹跳了。你得让水漂抬起一个角度入水，才能压出小冲浪，这个角度等于20度时你会像魔术师一样让水漂不停地弹跳。出手时让水漂旋转，它会飘得更直更稳。

水漂弹跳的秘密 冲浪腾空（图1.1）

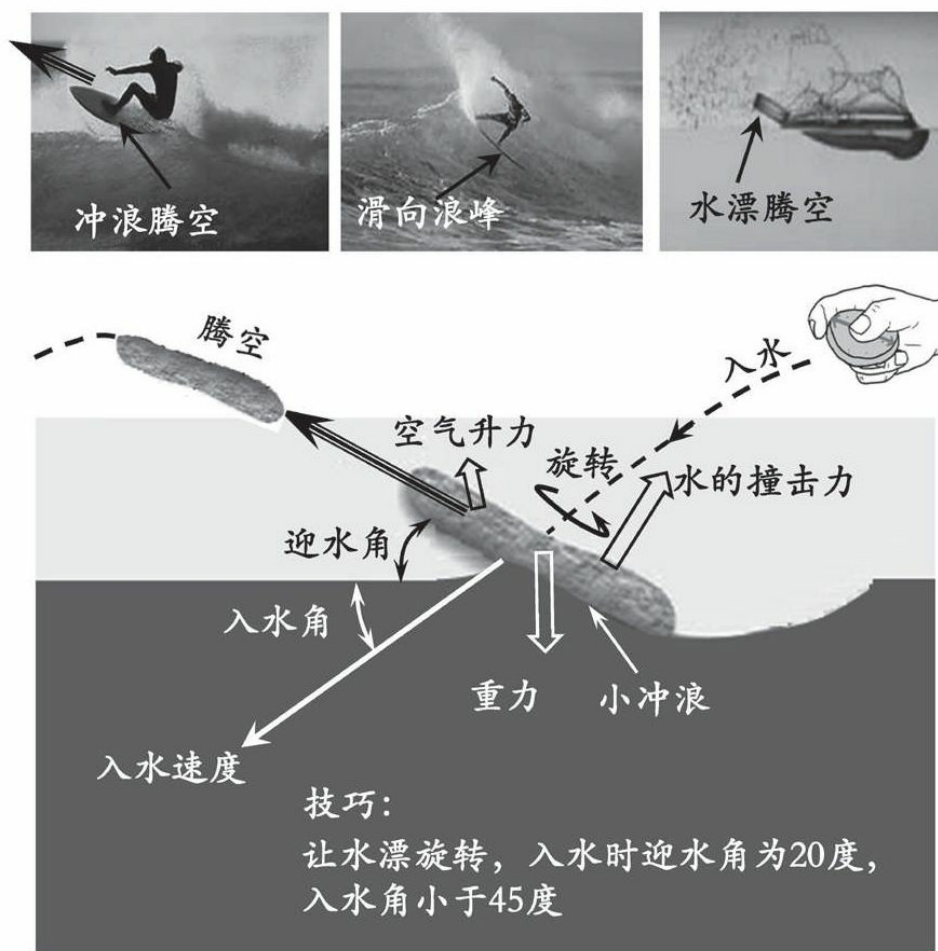


图1.1 水漂击水挤出波浪，做冲浪腾空

在池塘边，捡起一块扁平的石块，朝水面尽量水平地甩出去，使石块击水后，在水面连续弹跳，这就是打水漂。你可以把水漂为何能弹跳想象得极其复杂，甚至去看深奥难懂的科学报道。可是，既然那么容易弹跳，道理就应该十分简单。

大海边波涛滚滚，冲浪爱好者踏着冲浪板，获得速度后，像有速度的汽车冲上山坡一样，沿着波浪的斜坡面向浪尖滑行，离开浪尖时凭惯性腾空，再落到水面去追逐另外一个波浪。这叫冲浪运动，道理却和水漂弹跳十分接近。

水漂抬起适当的角度撞击水面时，将本来平静如镜的水面推出了一个小浪。小浪在水漂的腹部，或者说水漂骑在小浪上。于是，水漂沿着像斜坡一样的小浪，凭惯性就可以滑向空中，接着在重力作用下再次击水，进入下一轮弹跳。

如此，不难理解，水漂不能平着击水，那样产生不了有斜坡水面的小波浪。也不能垂直下砸，那样就没了向前滑行的惯性。

说得更深一点，水漂压水时推挤出有斜坡的浪，同时被水的撞击力消化掉入水时向下运动的那个速度分量。当这个向下运动分量完全消失后，就凭依然存在的水平速度惯性，在水漂尾部水压以及水漂头部一定的空气升力辅助下，如同冲浪板，沿着小浪的斜坡向上爬，尾部到达浪尖后凭惯性腾空。

水平速度分量要求足够大，腾空惯性才够。垂直速度分量要合适，才能有效压出斜波浪且不会突然沉下去。这就要求迎水角和入水角大小适中，否则要么水平速度不够，要么垂直速度不够或太高。

入水那一刻，水漂底面与水平面的夹角是迎水角。入水那一刻，飞行速度方向与水平面的夹角是入水角。每个角太大太小都不行。你可以分别想象是0度或90度这种极端情况时，看看会导致什么不利。于是一定有更合适的值。

迎水角和入水角加起来，就是水漂的底面与飞行方向的夹角，在航空上称为迎角（也称为攻角或冲角）。这个迎角使水漂能获得空气给它的一点升力（以后会介绍原因）。这个升力使水漂飞行时，不会下落太快，从而在弹跳帮助下飘得更远。迎水角和入水角到底多大合适？这个吸引了科学家的研究。

2004年，法国马赛大学的非平衡现象研究所的克拉尼特、赫森和伯克特三位科学家制作了一个带自动反射装置的打水漂机器，可以弹射出不同速度、转速、迎水角和入水角的水漂（铝制飞碟）。用高速摄像机拍摄水漂击水和弹跳过程，获得入水速度、迎水角、入水角和击水时间。一旦水漂碰水，可在水面停留20毫秒以上才弹起来，这个过程就在筑小冲浪并腾空。水漂入水速度不能太低，否则漂不起来。在他们的试验中，这个速度必须在每秒2.5米以上。

他们发现：入水角不能超过45度，如果超过45度，就会直接入水，不会有弹跳；迎水角为20度时效果最好，即不管其他条件如何，如果迎水角是20度，那么弹跳次数最多。这一发现在2004年1月1日发表在英国著名科学杂志《自然》第427卷第29页上。从此以后，20度这个迎水角，就被称为打水漂的魔幻角。

当然，还有一些其他技巧，可帮助弹跳次数更多。在此之前，先看看一些比赛结果以及比赛规则。只有在某种规则约束下，结果才有意义。

民间与正规打水漂 水漂鱼雷与打水漂飞行
(图1.2、图1.3)



图1.2 打水漂活动

以适当的方式扔向水面的石块，能在水面弹跳，普通老百姓偶尔也能发现存在这种现象。因此，民间一些人似乎从小就知道如何打水漂。

1989年由科尔曼·麦吉成立了北美打水漂协会并组织打水漂世界锦标赛。该项赛事按弹跳次数记名次。2013年9月诞生的一个记录是88次。1997年出现了苏格兰世界打水漂锦标赛，赛事规则为按石块漂浮距离记名次。其他一些赛事包括威尔斯公开赛和英国公开锦标赛。曾有过漂浮120米的纪录。要知道，掷标枪的世界纪录也就100米左右，手榴弹能扔出去40米都很难。如果向空中扔一块石头要落到120米远的地方，哪怕不考虑空气阻力等，从离地2米的高度以45度这样的能使自由落体飞得最远的角度出手，初速度至少需要达到34米/秒。让手甩出这么高的初速度，难度极大。可见，水上漂可以轻易让物体跑得更远。

宾夕法尼亚州的“河道水漂节锦标赛”采用的比赛规则是：水漂赛手自带一簌六个石片。石片必须是自然的，不能在石片上添加或剔除任何

有助于产生推进的东西。比赛之前，赛事裁判将检查水漂石片是否满足规则。对于特殊情况的裁定，裁判可以咨询宾夕法尼亚州水漂最高委员会。最高委员会的裁定具有终审效力。

比赛开始前，由某选手掷漂，第二次出现漂亮的弹跳时，比赛即可开始。每位选手有60秒完成一次掷漂。裁判将弹跳次数依次记录。

有趣的是，水漂运动涉及一些专用术语。这些术语的英文单词目前在普通字典里很难找到。依据掷水漂的水平和发挥，水漂可能会出现弹漂、爬漂、沉漂、空漂和危漂等情况。

打水漂的原理也启发了一些应用。巴恩斯（Barnes Wallis）在“二战”时期发明了水漂炸弹，1943年5月16日晚，英军用水漂炸弹空袭了德国鲁尔河上的三座大坝。炸弹在水面弹跳接近目标，降低了高空瞄准的要求。时速为390千米的轰炸机在距离大坝水平距离400米和垂直距离18米左右的高空，投下旋转炸弹。水漂弹在水面弹跳的道理与普通水漂差不多。

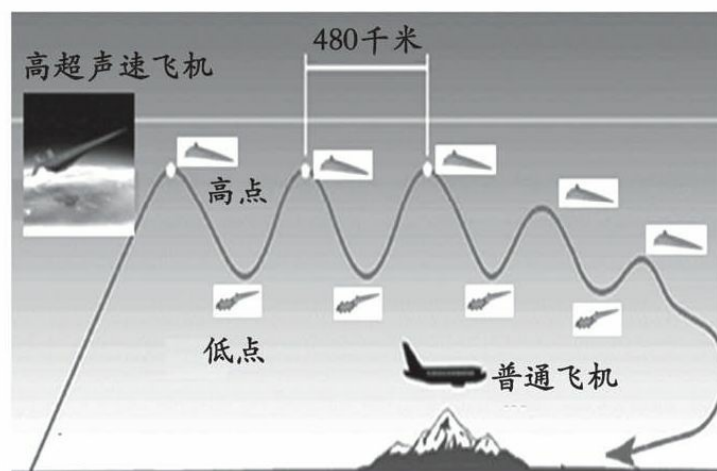


图1.3 “水漂式”飞行轨迹

普通飞机的飞行高度一般在20千米以下，利用稠密空气产生的升力飞行。卫星在100千米以上用轨道离心力飞行。在这两个高度范围的中间区域，两种飞行方式都不足以支撑有效飞行，但“水漂式”高超声速飞行可望解决问题，虽然“弹跳”方式和打水漂不完全一样。飞得低一点的时候，空气足够稠密从而发动机能点火产生推力，但长久平飞阻力很大；飞得高一点阻力小些，但空气很稀薄从而发动机无法点火。可以设想在35~60千米的高度以打水漂形式波浪式飞行。在“浪”底，由于空气稠密，可点燃发动机，产生的推力可使飞机爬飞到60千米高度。接近“浪”峰时，由于空气稀薄发动机只能熄火，但可凭爬飞阶段获得的惯性在阻力较低的“浪”尖高度滑翔一段距离。这样波浪式飞行，一个波浪就将近飞500千米的距离。

让水漂弹跳次数更多

我们希望弹跳次数越多越好，漂得越远越好。可是，我们普通人扔出去的漂，能弹跳10次就不错了。水漂在经历甩手、脱手、空中飞行、击水、弹跳和再入水的过程中，其实会涉及较为复杂的飞行原理以及与水的作用过程。史密斯就给出了一个打水漂方程式，来描述这些作用。

经过科学指导的打水漂则能将弹跳次数提高到数十次。例如，科尔曼·麦吉以12米/秒的速度和每秒14圈的转速扔出石块，弹跳次数达到了38次。

除了使迎水角接近20度、入水角尽量小、出手时让水漂旋转外，水漂的尺寸也有讲究。扁平石头厚度不超过6毫米，且直径在3~6厘米。石头的密度一般为每立方厘米2.5~3.3克，即每立方米2.5~3.3吨。如果取平均值，那么大致是2.9克/立方厘米。对于直径在5厘米，厚度为6

毫米的石片，体积大致是12立方厘米，重量大致是34克，即不到1两的重量。

如果观看北美河道水漂节比赛视频，我们会发现，赛手在扔水漂时，身体向后倾斜，手臂与身体呈45度左右。一般用大拇指与中指夹住石片，而用食指勾住石片外圆周。甩水漂时，出手高度离水面不要太高，以使入水角远小于45度；抬起石片，让其迎水角接近20度，同时尽量让石片高速旋转（在出手时，勾住外圆周的食指用力旋转）。

之所以要旋转，是因为石片飞出去后，姿态可能不平，可能是一侧先触水。这样，石片可能跑偏或者出现不必要的翻滚。如果高速旋转，就会产生旋转助稳效应（因为旋转也是一种惯性，稍微一点点扰动很难改变姿态），使飞行的石片保持基本稳定，使石片弹跳次数尽可能多。

如果弹跳次数多，漂行距离远，那么就可以欣赏到丰富的水面波动现象。在每个落水点，会产生一串串波纹，呈圆形向外扩展。这样一串串的由落水点发出的水波波纹，也称为涟漪。我们会发现一些奇妙的现象，虽然石片的形状不一定是圆的，但水面波纹是很圆的圆圈。前面时刻的落水点发出的圆圈扩散的速度比石片滑行速度慢，因此一系列落水点留下的涟漪，影响范围都在石片后面。但在末尾阶段，石片如果因减速出现爬漂，即慢慢地在水面漂，那么涟漪波纹可能跑得比石片快。这些水面现象，是极其丰富的水面波现象的一部分。

石片击水，一方面将水面击打出漏斗一样的压痕形成小冲浪，另一方面激发出水花。石片弹开后，周边的水回填，产生震荡，就激起了上面说的涟漪。水花中的水滴尺寸有大有小，小的近似呈球形。有种力量将小水珠包成圆球形，这个力量称为表面张力。

石片的飞行原理和旋转的稳定作用，水面波动现象，在后面会有更多的展示和讨论。

2. 不修边幅的飞石

如果你用形状一点儿也不扁平的石头打水漂，就没那么容易了。别说不能在水上漂，在空中也会惹恼空气遇到障碍。谁都知道，抛向空中的石头，在地球引力（也称为重力）作用下会加速下落。可是，除了重力，别忽略和小看那些看不见的空气分子。空气可不是那么好对付的。分子很小，可是它们多啊。它会挡在前面，你得推开它；它又在后面尾随，你得拖着它。就是不让你省力，它试图阻止小石头飞得更快、飞得更远。小石头目前还不了解空气的脾气与爱好，当然会遇阻啦。空气分子活蹦乱跳的，你在其中穿梭，不缠着你才怪，想想捅了马蜂窝是什么情况。

千奇百怪的下落与飞行现象（图1.4）

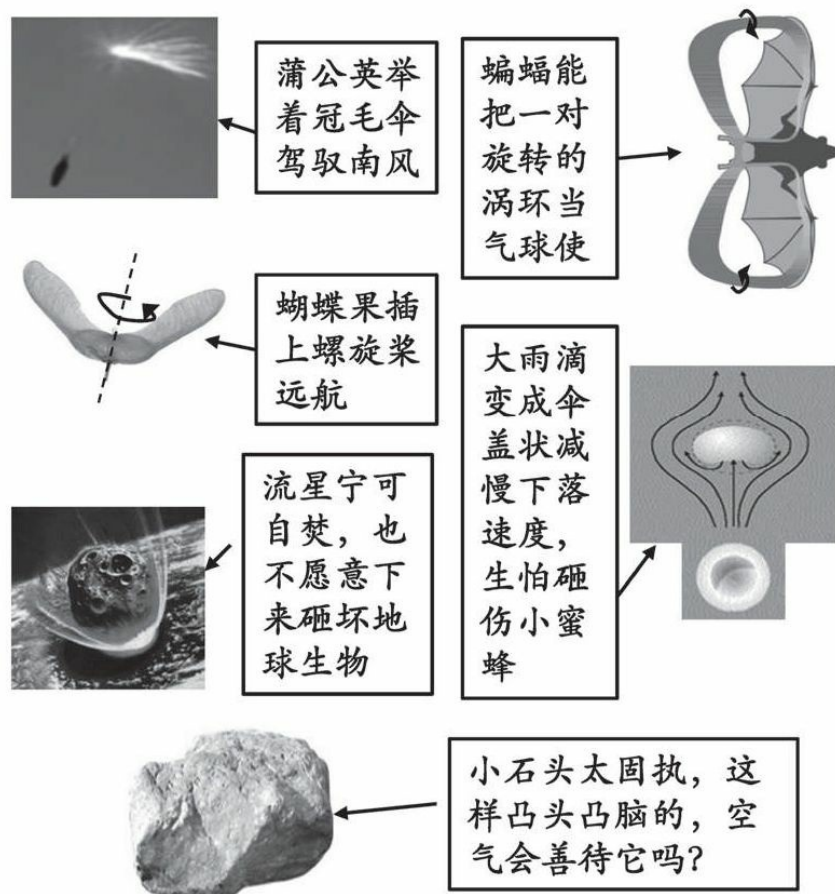


图1.4 各种形状的物体在空中

还是说石头，肯定令人失望：石头有什么好说的，无非就是说它飞着飞着会下落，空气还会使坏，让它飞得慢。可是别急，也不要下意识地对空气产生不好的印象。首先，即使这显得枯燥，那么能从枯燥中也看到乐趣，就不会在乐趣中看到枯燥。

这里之所以要说石头，是想说明这样一个道理：因为小石头不愿意了解空气的脾气，所以就会遇到困难。但不是所有东西都会像小石头这样不修边幅。就像人一样，修饰一下，别人可能对你更客气。空气不一定比人的境界高，难怪它也会以貌取人。蒲公英就知道举着华丽的冠毛伞，南风当然就成了它们的游乐场。大雨滴在空气中穿梭，会知趣地变

成伞盖状，这样空气就会帮它减小地球引力的加速作用，使雨滴徐徐下落，从而不会砸伤小蜜蜂遭遇官司。小鸟和昆虫选择了精美的翅膀，空气气流会帮助它们自由飞翔。飞机模仿了鸟，插上一对形体优美的机翼，空气便友好地在机翼上表演上吸下举的绝技，让飞机轻巧地在空中远航。空气让急速降落的流星带着一团火球、让旋转的足球走弧线、让美丽的雪花轻轻飘落。

但现在我们看看不修边幅的小石头会如何惹恼空气。

被小石头推挤和拖拽的空气（图1.5、图1.6）

我们耳听过风吹树叶沙沙响，也见识过肥皂泡泡或小气球炸裂时飞溅的水珠或小碎片。于是，我们不难理解，空气并不空，一定有看不见的东西，且有一定的内压。确实，大气中的空气并不空，由多得数也数不清的、小得看也看不见的、可以近似看成球的空气粒子（也称为空气分子）组成。暂时，我们不去看这些分子有多小、有多少和是什么。这里，我们只需要知道，它们是数也数不清的悬在空中到处乱跑的小球。

这种到处乱跑也称为分子热运动。每个分子跑的方向不一样，速度大小也不一样，各有各的自由。一不小心，一粒分子就会碰上另外一粒分子。正是这种热运动，我们才感知冷热，原来我们是被分子碰热的，否则将冻成冰块。正是这种相互碰撞，空气才有内压。

这是一块各个方向尺寸差不多的小石头。小石头在空气中飞行，迎面撞击空气分子，迫使挡路的空气让开。因此，挡路的空气会被推着走。人推着一群小孩往前走，就得使力。同理，石头推动空气，也给了

被推动的空气以力的作用。反过来，撞击石头空气也以牙还牙地给了石头迎风面的力，即石头的迎风面有了正压。这种以牙还牙，用物理学语言，就是所谓的牛顿作用力与反作用力定律。

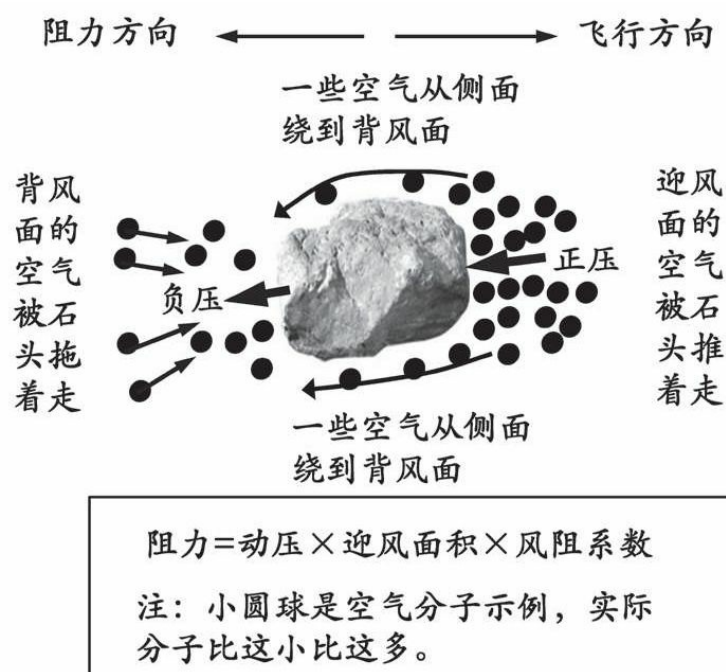


图1.5 小石头在空气中飞行

从10米高平抛半径5厘米的石头轨迹

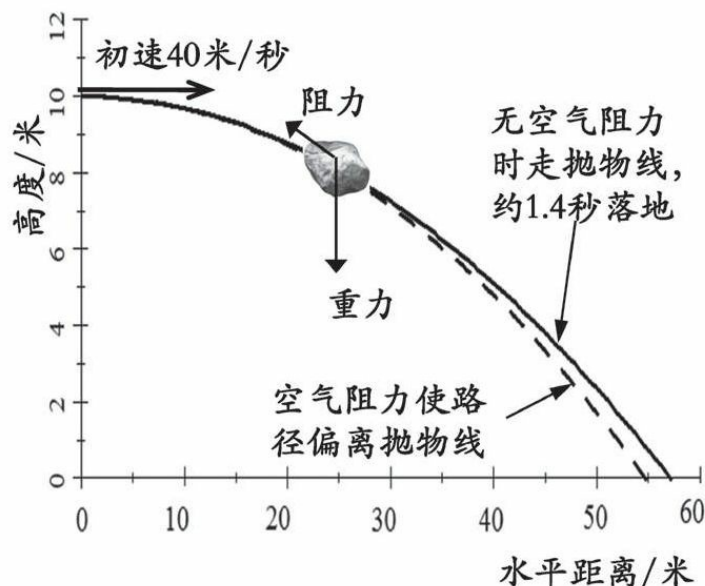


图1.6 小石头飞行轨迹

迎风面的反面是背风面。你拖着一群小孩往前走，需要使力。背风面的空气被石头拖着走，当然需要施加作用力，反过来被拖动的空气也就有一个拉扯石头的力，即石头的背风面出现负压。背风面的负压也可以这样理解，背风面的空气显然是被石头吸着走。只有负压才能被吸着走。你用一根管子对着水吸气，水就流进嘴里，表明吸气时，气压低了，水就顺着低气压进嘴里了。

当然，如果小石头沿着飞行方向比较细长，那么被推动的空气可以从侧面绕到背风面去帮忙。这里不考虑这种纤细的形状。从侧面绕到后面的空气太少，不起太大作用。背风面起作用的主要是被拖动的空气。背风面的空气又没长手去拉着石头，怎么会被拖着走？原来，如果不被拖着走，石头背风面就会试图被留下真空。所谓真空，就是真的很空，空得几乎什么都没有。由于空气分子有乱蹦乱跳的热运动，会去填满任何空隙，因此不会留下真空。

迎风面的正压与背风面的负压，均指向飞行反方向，就像你在人群中挤向前面去打架，劝架的在前面挡着你，从后面拖着你，于是你遇到阻力。石头迎风面的正压和背风面的负压，合起来就是一种阻力，称为压阻，指向飞行相反的方向。

你在地面拖动物体，地面给物体摩擦，即产生摩擦阻力。石头飞行时，其侧面与空气分子也会发生摩擦，因此也产生摩阻。

这种压阻与摩阻一起，形成总的阻力，指向飞行方向相反的方向，会将飞行的石头的速度降下来。因此，由于空气阻力的作用，飞行的石头比没有空气阻力的情况飞得更近。

小石头的飞行速度越大，与空气相互撞击的力度就大，因此阻力越大。这样形成的阻力，与飞行速度的平方近似成正比。也当然与空气的密度以及迎风面的面积成正比，这是因为空气越稠密以及迎风面积越大，参与作用的空气就越多。比例系数也称为风阻系数，或阻力系数。阻力系数与形状、姿态和速度都有点关系。物体越美丽光顺，越娇小纤细，阻力系数越小。凸头凸脑的小石头，阻力系数接近1。扁平石头阻力系数就小多了。因此，不知是否可以说，阻力系数就是美丽指数。美丽指数到底是多大，就是科学家去搞清楚的事情了。科学本身就是一种艺术，这么好玩的问题，在科学家面前当然能迎刃而解。

3. 谁都钟爱流线型 远走高飞的高尔夫球

具有流线型的物体和动物拥有的外形比小石头乖多了。现代飞机、高速列车和小汽车拥有美观的流线型取悦空气或水，使它们能顺畅地在空气或水中穿梭。流线型外形细长或扁平、表面形状光顺过渡、带有一定的弧度且表面没有坑坑洼洼。放在流水中，水能顺着这样的外形流

过，流过物体的水会贴着物面走，流水形态不乱。正因为如此，流线型的飞机可以飞得快、飞得远。没有体型，光有亮洁的皮肤也不一定行。高尔夫球就不是流线型，外表越光滑空气越添堵。它干脆让皮肤粗糙，把气流磕得乱蹿，这种乱蹿的气流反而帮忙消灭添堵的拐弯涡，使高尔夫球飞得更高更远。这可能预示某种道理，顺着来如果不行，那就反着来试试。

流线型，能顺利穿梭于水或空气中的外形（图1.7）

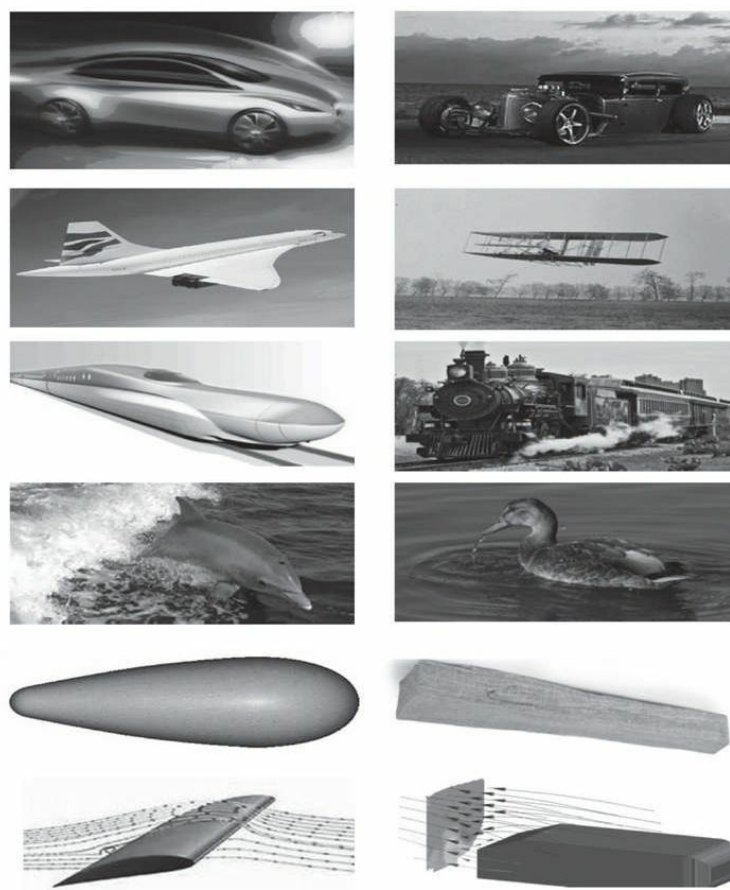


图1.7 流线型（左）与非流线型（右）

具有流线型外形的物体顺着长度方向运动时，水和空气更容易从侧面避让，一部分跑到背风面。这如同更瘦小的人或更细长的车，在稠密的人群中穿梭更顺畅一样。于是，需要被迎风面推着往前走的空气就少了，需要被背风面拖着走的空气就少了。不难理解，这种被推动和拖动导致的压阻就小多了。主要就剩下摩阻了。更形象地说，流线型外形纤细而又光顺，运动时对水或空气骚扰（扰动）很轻，水或空气也就不会给流线型物体施加太大的报复作用。

经过大风吹拂的山丘、经过雨水洗刷的山坡、经过流水冲刷的内河岸，形状也是某种形式的流线型。山形地貌有了这种形状，气流或水流就能顺利地流过，否则就会施加较大的作用，试图削平那些凹凹凸凸的部位。这有点儿枪打出头鸟的味道。大自然也懂得一点儿圆滑的道理。

于是，那些需要高速行进的汽车、飞机、列车等就会被设计成流线型。这种人造外形，与自然界经受气流或水流冲击形成的山形地貌或优胜劣汰进化出来的动物外形，具有异曲同工之妙。外形巨大的鲸鱼，在海水中能以每小时数十千米的速度游动，很大程度上得益于其拥有流线型外形。鲸鱼的速度大，躲避攻击或者捕获猎物的能力就强，因此更容易进化出巨大的体型。

物体走弧线的离心作用（图1.8）

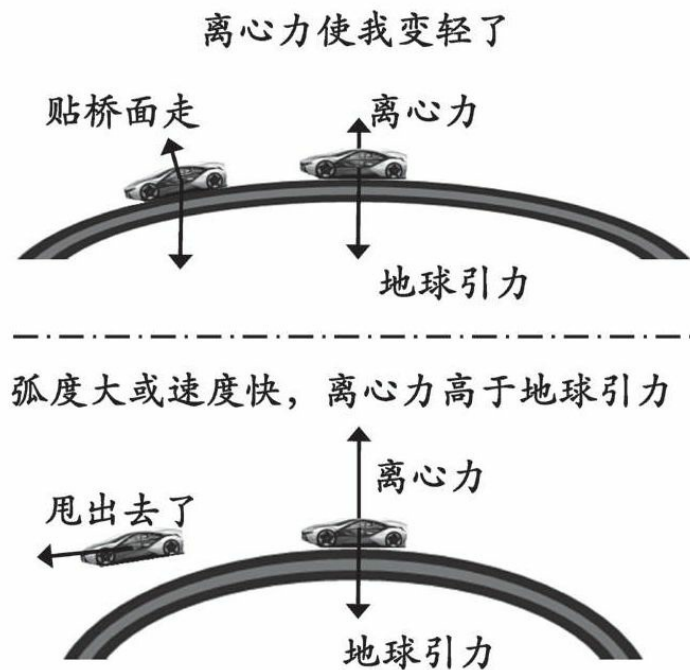


图1.8 汽车绕拱形路面行驶

既然是流线型，那么物体表面就带有一定的弧度。说到弧度，你可以先体会一下人行弧线的感受，尤其在游乐场。

回想一下在游乐场体验过山车、大转盘、大摆锤或者超级波波翻的滋味吧。尖叫声中，也许有人在喊：被捆住啦、加速啦、失重啦、感觉要甩出去啦、落下深渊啦、虚脱啦、实在是受不了。如果在快艇或冲锋舟上，风驰电掣，急行于波涛中，一跃踏上浪尖，再砰、砰，重重地拍下去，你可能会感觉船会散架或顷刻翻倒。

那种天旋地转的感觉，告诉你恐惧就是激情。如果说激情不恰当，那就说成刺激。据说这种“刺激”能增加皮质醇、咖啡因以及肾上腺素。在这种激情之中，人在被迫走弧线。我们所感觉到的失重或被甩出去，是因为被限制走弧线时，受到了一种指向弧线外侧的离心力的作用。这种离心力如果抵消重力，就让你有失重的感觉，特别是指离安全椅，就

让你有甩出去的感觉。以过山车为例，其运行路径是一圈一圈的，我们被捆绑在座椅上，时而进入内圈，时而贴着外圈。绕内圈时，感觉体重增加了，绕外圈时感觉体重减轻了，有被甩出去的感觉。这就是离心力的作用。

带弧线的流线型物体，在空气中（在水中也是如此）运动时，会迫使那些贴着表面走的空气走弧线。任何物体，包括一小团空气和一滴水，走弧线时，会受到一个离心力的作用。

如果要体验一下离心力与什么有关，那么可以试着用细绳拴一块小物体甩圆圈。此时，你感觉物体通过细绳在拉扯你，旋转线速度相同时，绳的长度越小，拉扯的力就越大。转速越快，拉扯的力也越大。这个拉扯的力就是离心力。如果你松手，物体就不绕你转了，就向远离转动中心（即你牵引细绳的手）的方向飞去。

如果需要说一点道理，可以这样理解。原来，运动方向也是一种惯性，你在维持物体旋转时，相当于在不断改变物体的运动方向。惯性本来使运动物体应该走直线，如果强迫其走弯径，就存在试图让其回归直线的离心效应，于是就得用一个力来抵消这个离心力，才能维持进一步走弧线。给运动物体一个力，力的方向垂直于运动方向，那么就不会改变速度大小，而改变速度方向。

对于有弧线表面的物体，近似看成一个圆弧，就会对应一个圆的半径，称为曲率半径。这个半径与上面所说的甩圆圈的细绳长类似，于是，流体绕弧线表面流过时，如果曲率半径越小，即拐弯越急，离心力就越大，越容易甩离物面。因此，为了使气团或者水团能贴着物体走，流线型物体的曲率半径要足够大，也就是说不要弯曲得太厉害。

设想你在公路上驾驶一辆小汽车，要横穿一条小河，小河上有一弧形拱桥，直通河对岸，从而你开车过去时需要在这个数米长的弧形拱桥上走。拱桥如果不高，那么小汽车就能平稳地贴着拱桥表面开过去。如果车速足够快，那么我们会感觉体重减轻了一点（即臀部与坐垫之间的压感小些了），即有失重的感觉。驾驶小汽车经过上面所说的拱桥，就受到了离心力，离心力抵消了一部分地球引力，使在车里的你感觉自己轻了些。

小汽车行驶在拱桥上，如果拱桥太高使弧度太大，那么速度足够快的小汽车就可能从拱桥最高点附近飞离。就是说，拐弯太急，离心力的作用就把汽车甩出去了，即汽车与路面分离了。

非流线型 拐弯涡（图1.9）

流线型物体虽然有弧线，但正是因为弧线，可以做成没有拐弯太急的位置，离心效应不会太大，气流就会贴着物体走。

反之，如果是非流线型，从迎风面避让绕到背风面的气流，需要拐弯就太大或者太急。在拐弯太急的地方会在某点甩出去，这种现象在专业上称为气流分离。如果在水中运动，这些现象就是水流分离。

分离使从迎风面绕到背风面的空气就少了，因此主要是背风面下游的空气被物体拖着走，需要的力就大，因此阻力更大了。分离出去的气流与背风面被拖进来的气流方向相反，因此卷曲成旋涡，称为拐弯涡（专业上叫分离涡）。

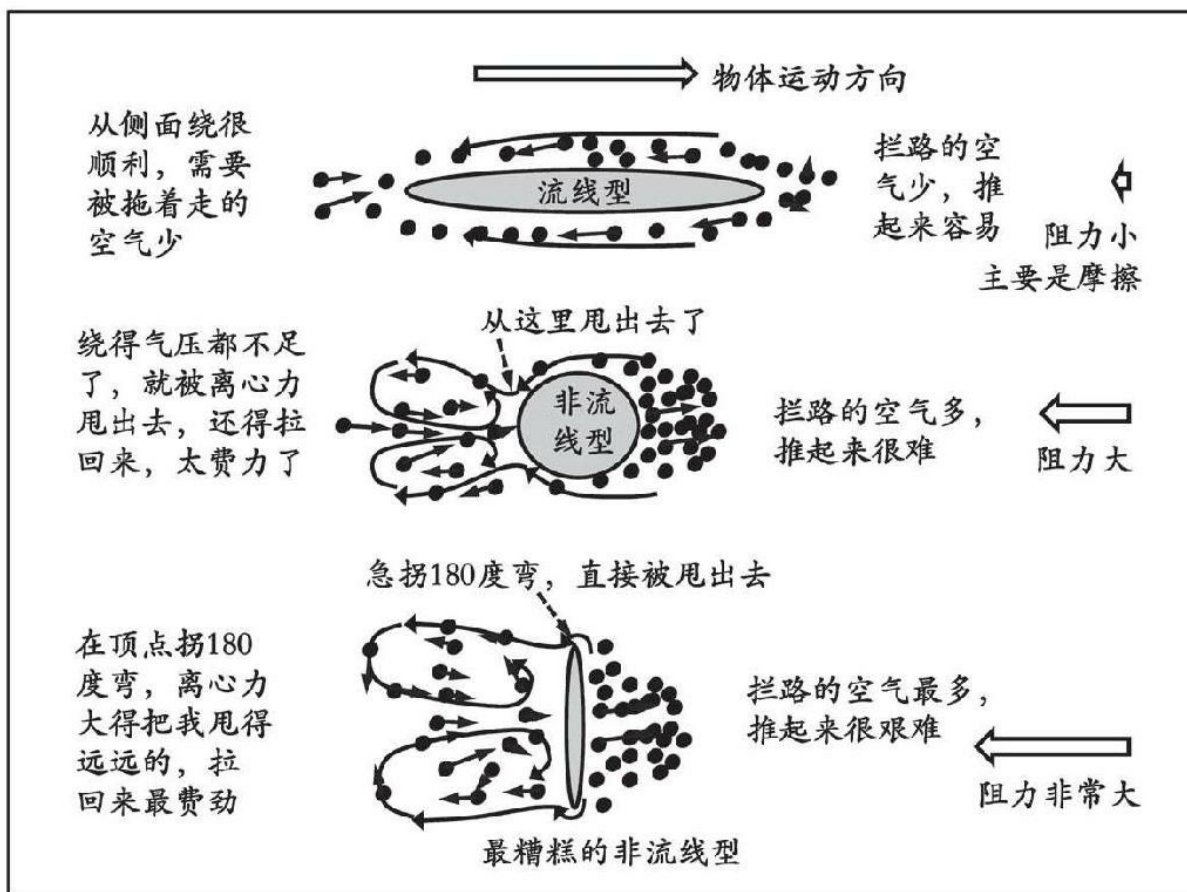


图1.9 流线型与非流线型的气流与阻力

甩出去的空气被拐弯涡卷进背风面，绕了一大圈。人走路时，错过了该去的地方，绕一大圈折回去时，已经没什么力气了。类似地，对于非流线型外形，出现拐弯涡后，背风面气压就不足了，相比于大气原有气压，就会有较大的负压。于是总的压阻就大。

让流线型物体的长度方向朝着飞行方向，如扁平石头顺着飞行方向，那么就很难出现分离，压阻就小。如果长度方向垂直于风向，那么阻力就大。你抛椭圆形状的橄榄球，比抛呈球形的排球更远，就是因为橄榄球比圆球更接近流线型。

即使外形像流线型，如果表面不光滑，那么也不是流线型。在地上

拖东西，如果地面凹凸不平，那么摩擦阻力就大。如果地面平坦，摩擦阻力就小些。一般情况下，越光滑，摩擦阻力越小。物体在空气中或水中运动，也会受到摩擦阻力。越光滑，摩擦阻力就会越小。

非流线型 高尔夫球的凹槽（图1.10）

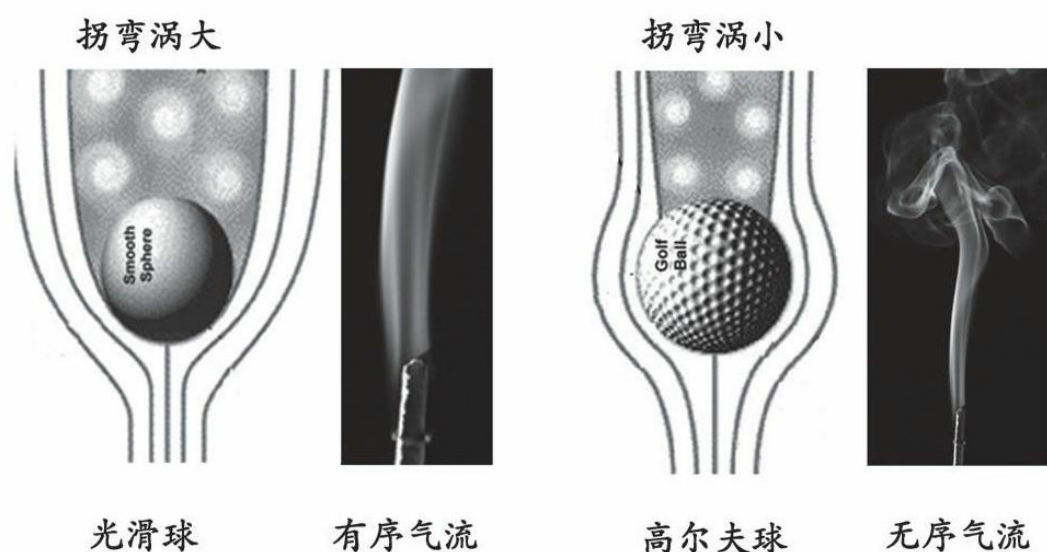


图1.10 光滑球与高尔夫球

但高尔夫球用光洁的外表反而没用，因为它是非流线型外形。为了减小摩擦阻力，我们想当然应使高尔夫球表面尽量光滑。可是，不是流线型的高尔夫球百依百顺也没有用，光滑反而不能取悦空气，阻力会更大。那干脆软的不行来硬的，让表面粗糙来死磕空气，看你空气怕不怕。还真怕，空气让步了，阻力减小了。当然，咱们不能这样用吓唬人的思维讲道理。既然有结果，那就一定有理性的原因。苏格拉底的因果定律说，每一个结果都有特定的原因或者多个原因。

重量不能大于45.93克、直径不能小于42.687毫米的高尔夫球，有酒

窝型凹坑时，可以飞得更远、更稳。这是偶尔发现的奥秘，事先并不知道为什么。原来，一百多年前，英国工程师韦林·泰勒意外地发现：用过的高尔夫球的表面会出现不规则的破损，且这些旧的高尔夫球比表面光滑的新球飞得更远。他立即做了许多实验，设计出了带酒窝凹槽的高尔夫球，这样的球飞得更远。

带了酒窝，就更偏离流线型了，摩擦阻力应该更大，按理飞得更近了。为何反而飞得更远了，即阻力更小了？

原来，对于非流线型球状物体，很容易提早产生拐弯涡，产生很大的压阻。如果带有凹槽，就很容易把气流磕乱，形成所谓的湍流，让气团找不到方向。坐飞机时，偶尔出现空乘人员提示：飞机遇到一些湍流，有些颠簸，请系好安全带。可见，湍流能乱得使飞机都颠簸起来。我们常见的香烟柱，下面一段是直的烟雾，上面一段是乱的烟雾。乱的部分就是湍流，烟丝一会儿串到这里，一会儿串到那里。烟雾萦绕即反复缠绕，就是对湍流的一种刻画。

行走方向乱的幼儿，更容易钻到别人家的房间。被打乱的气流无规则地到处乱跑，很容易钻进分离留下的空隙，使一旦有分离就可能被打乱的气流填进去。因此，拐弯涡就不容易过早发生。于是本来比摩阻大得多的压阻降低了，飞得更远了。

高尔夫球的凹槽显然不是随意刻的，只有满足一定的规律才有效。六角形的凹槽最有效。现在可以先思考一下，为何六角形最有效，是否与六角形蜂巢和六角形雪花的形成有相似道理？

更令人纳闷的是，烟柱到了一定的高度怎么会变乱，哪怕在安静的室内。如果直直的升起，烟柱就局限在一个窄的上升通道中。也许只有

变乱，上下左右前后乱蹿，才可以拓宽上升的通道。变成湍流可以快速拓宽活动的范围，谁都不情愿憋在窄细空间中受约束。于是不难理解为何会出现湍流。

4. 省力的橹与飞行的秘密 迎角的升力与抬头效应

我们都有这样的常识，飞机与鸟之所以能在空中飞行，是因为飞行的机翼与翅膀能产生升力。可是，很少人知道，机翼和翅膀产生升力的方式或者说原理，居然和公元前汉代发明的橹在摇动时产生推进力的原理完全一样。不知是否可以说，是汉代的人最早掌握了飞行的本质原理。据说受橹的启发，英国人发明了螺旋桨，美国人改进了游泳运动员的姿势。这个秘密就在迎角，即扁平的物面与运动方向有夹角。迎角使扁平物体的一面被水流（或气流）撞击，一面被水（或空气）拉扯，产生推进力或升力。

船橹的秘密（图1.11、图1.12）

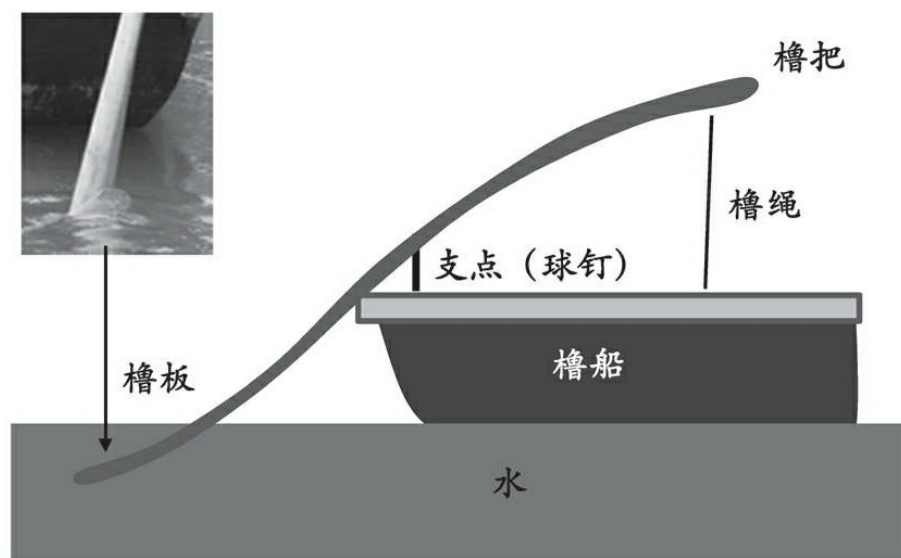


图1.11 橹的基本构造

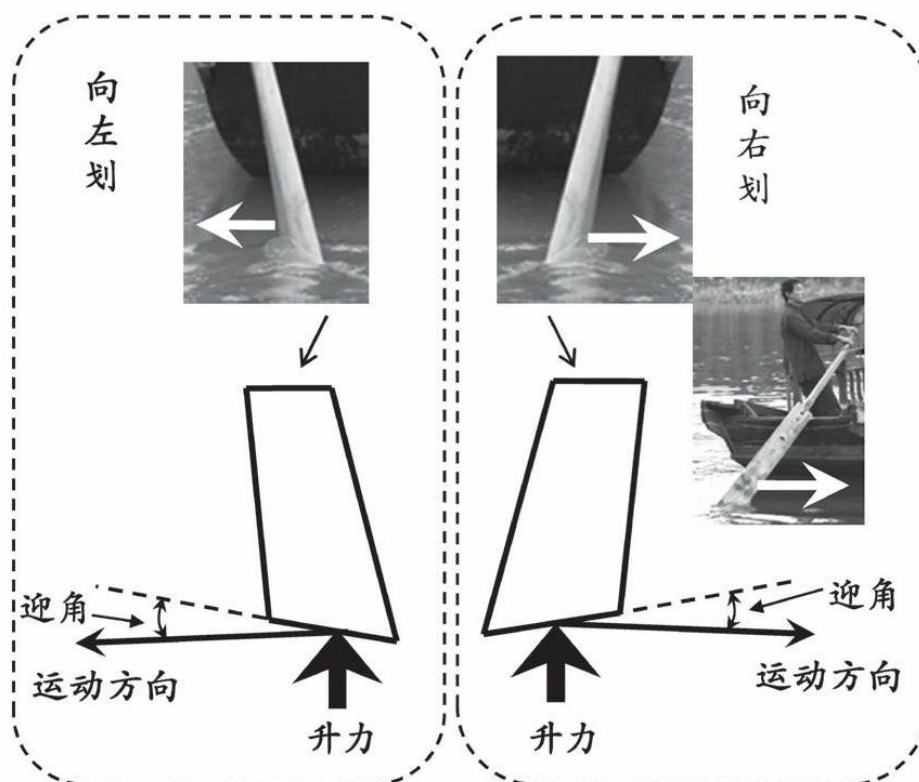


图1.12 橹板的交替运动产生升力即推进力

宋代慧远禅师的《点绛唇》曰：“来往烟波，此生自号西湖长。轻风小桨，荡出芦花港。得意高歌，夜静声偏朗，无人赏。自家拍掌，唱彻千山响。”可见一边划船还能一边唱歌欣赏美景，这说明划船的桨叶还是比较省力的。

船夫有三篙不抵一棹、一橹抵三棹之说。可见，橹的效率最高。人们可能认为橹利用了仿生学原理，仿照了鱼尾，通过像鱼摆尾一样摇摆橹来产生推动船前行的力。事实上，橹不是像鱼摆尾那样通过摇摆获得推力，而是利用了神奇的迎角效应，与机翼产生升力的道理一致。

东汉刘熙在《释名》一书中就提到了橹，尤其指出橹在摆动时能产生力，且力沿着船脊方向。船橹一般通过支点固定在船尾，握手的橹把则通过橹绳与船缘链接，可以在不用的时候稳定船橹并配合橹手左右摇动橹把。摇橹手面朝船的一侧，向船的两侧方向摇动船橹，使在水中的橹板朝船的两侧来回划水。橹手通过转动橹柄，使橹板在来回划水时，橹板的一面总是与运动方向（即橹板划水时向两侧的运动方向）有一夹角。

这个夹角也叫迎角或攻角。迎角效应使橹板产生了垂直于橹板朝两侧运动的方向的升力，也就是推动橹船的力。往下我们会用带迎角的小石片来说明为何会产生升力。

当然，除了与运动方向垂直的升力，也存在与运动方向相反的阻力，尤其是摩擦阻力。但迎角适当的话，升力可以比阻力大10倍左右。橹手摇橹时，主要朝（橹板在水下运动的方向）相反的方向发力，这样橹手就只需要克服橹板较小的阻力，却在橹板上产生了较大升力。另外，这个升力方向指向船前进的方向，对船的行驶而言，这就是推力。

因此，摇橹手只需要付出较小的体力，就可以产生较大的推力。这就是橹省力的原因。产生升力的迎角效应有一定的普适性，下面用飞行的石片做通俗的解释。

迎角的升力效应（图1.13）

设想有一块扁平的石块，某时刻，其飞行方向与石片有一夹角，即上面所说的迎角。假定迎角不太大，比如说10度左右。

飞行时迎着飞行方向的那一面称为迎风面，迎面撞击空气，当然会产生正压，于是迎风面也可以称为压力面。另一面需要拖着空气走，因此很难产生正压，一般会产生负压，故称为吸力面。这个道理与前面介绍的非扁平石头迎风面产生正压和背风面产生负压的道理是相似的。

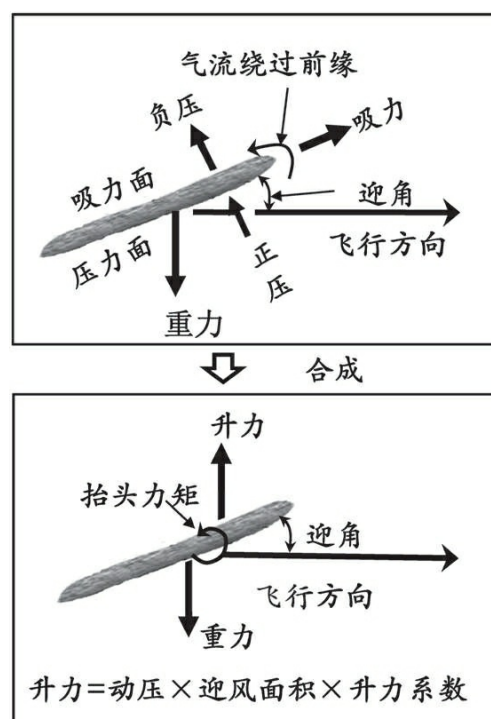


图1.13 迎角产生升力和抬头力矩

可是，如果仅有迎风面的正压和背风面的负压，那么正压与负压的合力就垂直于石片，而不是垂直于飞行方向。牛顿就认为石片受的合力垂直于石片。实际上，这种气压差带来的合力近似垂直于飞行方向，是与重力方向相反的升力。

原来，飞行的石片迎面撞击空气时，前方被推动的空气有一部分绕

过前缘到了背风面。绕过去时要拐很大的弯，从而感受到较大的指离前缘的离心力。这个离心力抵消了一部分当地气压，因此给前缘施加负压，产生吸力。前缘吸力指离前缘，与迎风面的正压与背风面的负压形成的合力，令人不可思议地近似地垂直于飞行方向，即这种合力就是升力。

有一点就想不明白了，如果前缘很尖细，那么提供的吸力能大到哪里去？原来，前缘越尖细，那么拐外越急，离心力就越大。因此，合起来的吸力就那么大。问题是，迎风面、背风面和前缘产生的合力为何偏偏与运动方向垂直？道理是这样的，既然合力不会垂直于石片了，那垂直于什么，才能用最简单的语言可以表述？自然规律追求简单、简约、简美，如果什么语言表述最简练最可理解，那么规律就能用这个语言表述。想来想去，没有比垂直于运动方向这种描述更省劲了。自然就说，那就垂直于运动方向吧。于是，这个力就真的垂直于运动方向。

当然，物理学家能用物理学满足的基本规律（如牛顿定律）并借用数学推演，证明这个结果。我们既然能以刚才懒惰的逻辑来指明结果应该是那样，何必在这里翻开连科学家都费劲才能搞清楚的推演？

如果迎角适合，那么石片受到的阻力（指与飞行方向相反的力）主要只有摩擦阻力。上面的气压差主要贡献升力。

这就是迎角产生升力的本质原因。飞行的石片、船的橹板、推进用的螺旋桨、飞机的机翼、鸟的翅膀、喷气发动机的叶片，甚至飘落的树叶，都是靠这种迎角产生气压或水压差，形成所需要的合力。对于飞行，这个压差形成的合力是升力。对于推进的叶片，这个力是推进力。对于舵面，这个力就是用于操纵方向的力。

适当的迎角能产生升力，但迎角太大则物极必反。最糟糕的情况就是横着飞，即迎风面与飞行方向垂直，此时如同前面提到的石头，只有阻力没有升力了。于是必然有一个最佳迎角，使升力最大。鸟翅膀的最佳迎角一般在12度左右，一般飞机机翼的迎角也接近这个数。

当然，石片边缘的其他部分不会像前缘这样提供恰到好处的吸力，使真实的气压差的合力也贡献一部分阻力，但这不影响对迎角产生升力的本质理解。

迎角的抬头效应

我们都知道如何更轻易地去掀翻或推倒一个物体，这比移动一个物体更容易。我们熟悉的杠杆原理就是一种四两拨千斤的方法。用专业一点的语言，就是力矩效应可以使物体改变姿态。你去推一件物体的重心，物体会移动但不倒。如果你在偏离物体重心的位置推物体，物体就转动或改变姿态。推力偏离重心越远、越容易改变姿态。推磨盘就是这个原理。

机翼或小石片在气流中带迎角运动，产生的升力一般不会作用在它们的重心上。由于是机翼的前缘主动去撞击前方的空气，因此迎角效应集中在离前缘不远的地方，于是产生的升力作用在离前缘不远的地方，即一般在中心点（即重心）与前缘连线的某位置上。由于升力在重心之前，与重力不作用在一个点上，而是作用在试图迫使机翼抬头的点上。于是，迎角效应产生了所谓的抬头力矩。飞行的石片如此，飞机的机翼和发动机的叶片也是如此。

如果你扔一块石片，那么飞行过程中稍有偏转，就会产生由迎角效

应带来的抬头力矩，于是石片可能改变姿态甚至翻滚。为了姿态更稳定，可以在将石片甩出去的时候，用手指带动石片旋转，看看旋转的石片是否飞得更稳定。

翼尖涡与桨叶涡（图1.14）

通过带迎角来获得升力的翅膀会在下表面产生正压，上表面产生负压。这种正负压差居然导致在翅膀两侧的翼尖附近出现一对往下游延伸的、旋转方向相反的旋涡。

划船桨时，在桨尖也会产生一对旋涡，道理是一样的。

原来，机翼下表面的气压比机翼外侧的环境气压高，于是会向翅膀外侧吹气。机翼上表面的气压比机翼外侧环境气压低，于是会向翅膀内侧吹气。



图1.14 翼尖涡与桨叶涡

这样形成的吹气方向相反的气流，到了机翼下游直接接触后，就变成了气流相互搓动。不同方向的气流相互搓动，就会卷起旋涡，就像你用两个手掌通过向相反方向运动搓擗面杖，擗面杖会旋转一样。你双手抛篮球时，往相反方向转一下掌，篮球就会转起来，道理是一样的。其

实，你单手也都能这样将篮球转起来，只是双手更容易做到。

机翼每一侧的翼尖都有同样的效应，因此就会出现一对旋转方向相反的旋涡。这等价于说，由于机翼上表面是低压，机翼下方的高压驱动气体翻到上面去，于是就形成了翻转涡。这与拐弯涡不同。

对于飞机，这对翼尖涡会将云层卷起，在飞机下游很远的地方都能看到。由于旋涡来自于机翼上下表面气压差，因此越大的飞机，升力越大或者说机翼上下表面的气压差越大，翼尖旋涡就越大越强。

5. 带着气流旋转 弧线球

运动的机翼靠抬头获得迎角后产生升力。抬头是相对的，若让空气气流低头，效果应该是一样。物体旋转能带动气流旋转，等价地产生气流低头或抬头效应，产生类似于升力的力。由于这个力不一定用于让飞机升起来，但与飞行方向垂直，因此广义地称为侧向力。于是不难理解，足球、篮球和乒乓球如果旋转，就会被侧向力驱动它们走弧线。

侧向力的产生 马格劳斯效应（图1.15）

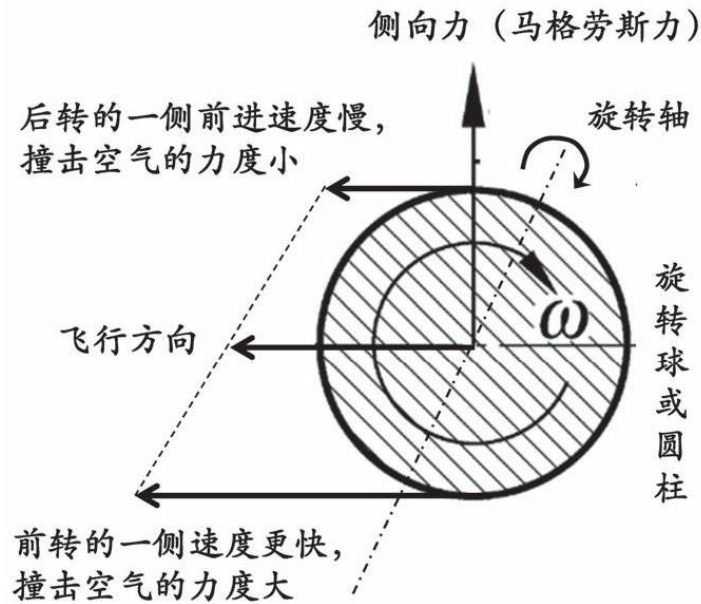


图1.15 旋转飞行的物体产生侧向力

物体如果一边穿梭一边旋转，在推挤空气时也把空气带着旋转起来。向前转的一侧（称为前转面）与向后转的一侧（后转面），撞击空气的速度从而力度不一样。力度大的那一侧会受到空气更大的气压作用，于是就产生了从前转面指向后转面的力。这个力与运动方向垂直，因此称为侧向力。这种效应是马格劳斯发现的，因此称为马格劳斯效应。产生的侧向力称为马格劳斯力。

说得更细致点，在前进一侧，旋转导致的物体表面线速度与飞行速度具有相同方向，带动空气以更快速度撞击前方空气，形成更高的正压。在后撤一侧，旋转导致的物体线速度与飞行速度方向相反，带动的空气速度低，撞击前方空气的力度小，因此正压低一些甚至形成负压。合起来就形成了前进侧指向后撤侧方向的力。显然，旋转速度越快，前进一侧与后撤一侧的速度差就越大。因此，产生的侧向力与旋转角速度成正比。这是一种理解方式。如果你借用前面对迎角的升力效应的理

解，也可以认为物体旋转时相当于迫使前方的气流低头或者抬头（是逆时针转还是顺时针转），这与物体抬头或者低头产生的效应当然类似。

由于侧向力与运动方向垂直，因此会迫使旋转的物体走弧线。

球类运动中的应用 弧线球与电梯球（图1.16）

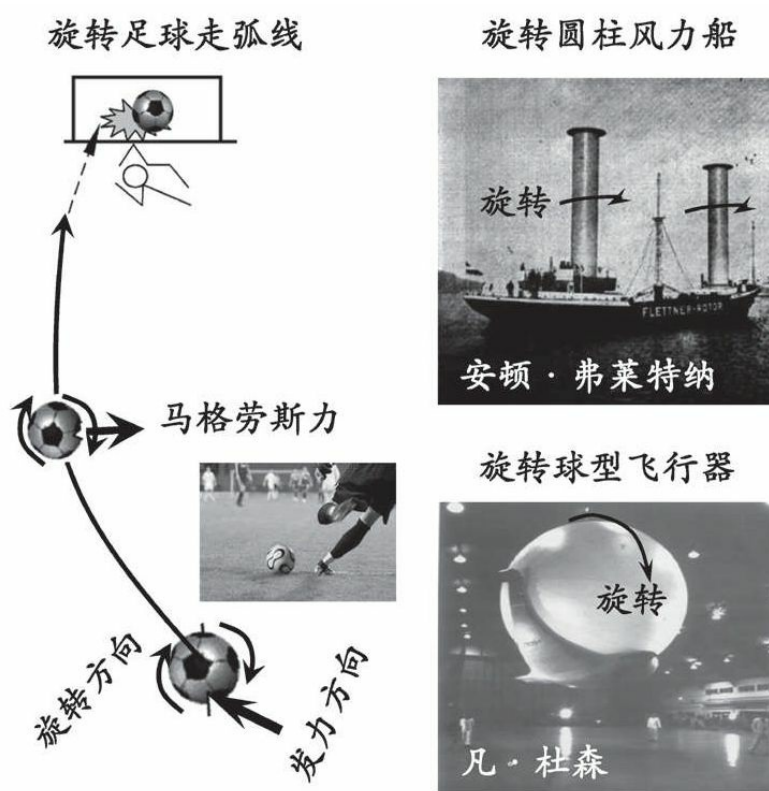


图1.16 马格劳斯效应的应用

20世纪50年代，日本乒乓球运动员利用马格劳斯效应发明了弧线球发球技术，发球时让球旋转，球就走弧线，以此压低或改变线路，同时让对手去接旋转着的乒乓球，增加接球难度。高尔夫球如果旋转也会偏离它的行进方向，使落点远离洞口（马格劳斯效应导致所谓的Slice球或

Hook球）。棒球和板球运动员也会投出各种弧线球，给对方击球手制造麻烦。于是，在球类竞技体育中，马格劳斯效应随处可见。

以乒乓球为例（过去直径为38毫米，质量为2.5克，后来改为40毫米和2.7克），如果用球拍搓乒乓球下表面使其产生下旋（即靠近身体的部位向下旋转），那么前飞的乒乓球就产生向上的侧向力，如同升力。这时乒乓球就会飘得更高。如果搓乒乓球的上表面产生上旋，那么就产生指向下方的侧向力，乒乓球会绕着低矮的弧线贴着球台走。

在足球运动中，高水平的职业球员利用马格劳斯效应，在罚点球或者任意球时通过让足球旋转，踢出漂亮的弧线球。如果右脚踢球使球的近脚一侧朝左旋转，那么产生的侧向力就指向右边，球会向右走弧线。当然，也可以让足球上旋、下旋和右旋，不难想像足球会分别走低矮弧线、腾高弧线和左侧弧线。发定位球时，要想绕过人墙入网，恐怕最好的办法是依靠让足球旋转起来的马格劳斯效应了。对于直径为22厘米和质量为0.430千克的足球（不同时代的足球大小和质量有区别），若从距球门25米处罚踢任意球，且以每秒25米的速度踢出，同时让足球每秒转10圈，则由马格劳斯效应产生的侧向力可以估算为4牛顿左右，足以让足球在入门前横移4.5米左右的距离，此时，守门员扑球难度增大。

神秘的电梯球，在接近球网时，突然下掉，让守门员猝不及防。这可能是向下的马格劳斯力突然增大了或者向上的马格劳斯力突然变小了或者反向了。因为马格劳斯效应也与飞行速度有关。速度大的时候气流可能被磕出湍流，和高尔夫球靠凹槽磕出湍流是类似的，于是足球背风面不容易产生拐弯涡。此时阻力小，足球只会慢慢减速。但足球速度低到一定程度后，磕不出湍流了，拐弯涡趁机冒出，用阻力强拽足球，让其速度迅速下降。此时，与速度成正比的马格劳斯力突然变小。如果足球的旋转事先使马格劳斯力朝上，速度突然变小不使足球突然下掉才

怪。

驱动航行的马格劳斯力

1924年，德国航空工程师安顿·弗莱特纳在船上树立一根可以旋转的圆柱，风吹圆柱时，因马格劳斯效应产生了垂直于风向的侧向力，可以牵引船的航行。这就是一种基于马格劳斯效应的风力推进船。

20世纪30年代初期，美国人尝试过用旋转的圆筒替代常规机翼来产生升力。1978年加拿大人凡·杜森用高强度的弹性材料卡芙拉制作一个大浮力球，用一根水平轴穿过球体的两极，水平轴的两端分别装有发动机，巡航飞行时带动球体旋转，产生的马格劳斯力作为升力。后来，有人成立了马格劳斯宇航公司，专门研制和生产基于马格劳斯效应的载人飞行器。

1.2 失稳、旋转与振动让世界更灿烂

万物都在运动，旋转与振动也是运动的形式。有的运动来自于外力，有的来自于自身不稳定。它们丰富了运动的自由度，让物体的能量可以存储与转换，恰如我们生活中的储蓄与消费。空气中氧气分子和氮气分子的热运动除了移动，还有转动。这种分子很怪，一个方向的移动分得了多少能量，那么绕一个轴的转动也得分多少能量。你搞不清它们为何要等贵贱均平富。如果空气分子没有转动，那么大气温度不知要高多少，转动分走了一部分吸收的太阳能。旋转蓄含了一种力量，物体遇到麻烦侧歪时，这种力量就会出来帮忙。旋转使相邻的星星不会受制于万有引力而撞在一起。振动让世界拥有节奏和旋律。如果没有振动，我们就没有时钟，也没有动听的音乐。旋转会导致物体与空气共舞，否则旋转的球类不会走弧线。自发的旋转往往来自于失去稳定，但主动旋转又会使物体更稳定。正是因为有旋转，才有我们的灿烂宇宙，才让地球不即不离恰到好处地靠近太阳，才让宇宙天体不会撞在一起。可以说，是因为有了旋转，才有地球生命。那就让我们在失稳、振动和旋转的世界甚至宇宙中走一遭吧。

1. 失稳中的多姿多彩或险象重重

沉稳就是一种情绪的稳定状态，失控或者冲动就是一种情绪失稳状态。下雨天在路面滑倒就是一种失去稳定的结果。失去稳定可能导致危害，也可能有利用价值。失去稳定可能演化出多姿多彩的有序现象，也可能演化出捉摸不透的混乱形态。有的影响小，比如骑自行车滑倒的影响范围极小。有的可能大，比如传说中的蝴蝶效应。有的系统本身不稳定，为此人们发明了巧妙的使系统稳定的方法，使本身不稳定的标枪、

弓箭、飞机与游船等变得稳定可控。稳定往往比追求变化更重要，至少是追求和谐变化的基础。一个不稳定状态，失稳后可能演化出一个稳定状态，“野渡无人舟自横”就是这种情况。

野渡无人舟自横（图1.17）

一艘自由的船在水流中也会像小石片或翅膀在气流中一样，其竖直方向与水流方向之间稍有偏转，就产生迎角效应，包括升力效应和抬头效应。对于船这样的外形，迎角效应虽然没有机翼那么大，但足够可以改变姿态。

野渡无人舟自横就是这种原因导致的，该诗句出自《唐·韦应物·滁州西涧》：“独怜幽草涧边生，上有黄鹂深树鸣。春潮带雨晚来急，野渡无人舟自横”。

没有人操控的船，如果竖直对着水流方向，那么稍有偏转，就产生上面所说的令其进一步偏转的迎角效应。因此，船身竖对着水流方向，是不稳定的，会慢慢偏离方向。

船横对着水流方向则不会失稳。即使遇到偏转，那么试图偏向上游的一侧撞击水流的力度就大，产生的额外阻力作用在这一侧，就会阻止

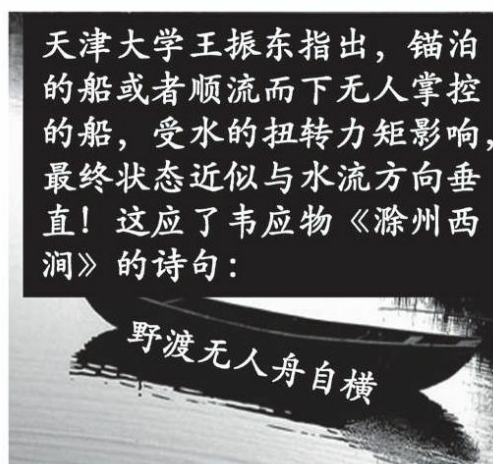


图1.17 野渡无人舟自横

其进一步偏转。试图偏向下游的一侧撞击水流的力度就小，减小那侧本来的阻力，因此也会阻止进一步偏转。

因此，横对着河道（即流水方向）的无控船只更稳定，即使遇到扰动偏转，也会产生使其恢复到横向的力。这就是野渡无人舟自横的原因。

秋风扫落叶与飘落的纸片（图1.18）

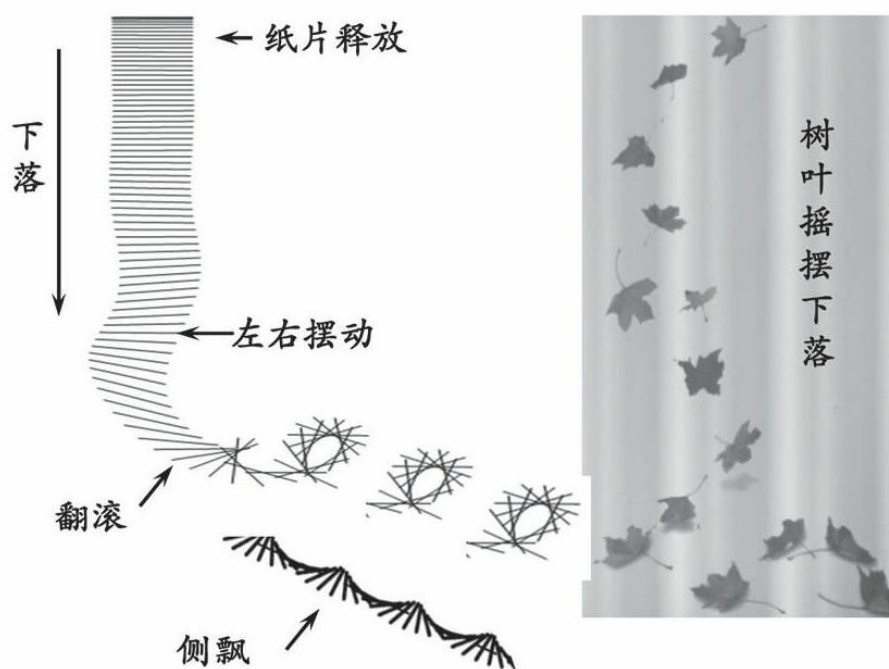


图1.18 纸片、树叶等下落和飞行的混合状态

树叶下落时可能会翻滚以及侧飘。这些现象淋漓尽致地出现在一些诗文中，例如《隋·孔绍安·落叶》：“早秋惊落叶，飘零似客心。翻飞未肯下，犹言惜故林”。还有司空曙的“随风偏可羨，得到洛阳宫”。我们看到的大多数树叶落下时，要么摇摆，要么翻滚。在垂直风的带动下，

落下的树叶还有可能飘起来。树叶形状比较复杂，经常有五个角。如果是长方形名片，则更好理解失稳和翻滚，虽然道理和树叶是一样的。

现在用名片或扑克牌等纸片来做一个实验。端平长方形纸片轻轻松开，让它平躺着开始下落。重一点的纸片会一直平躺着直接落入地上。轻一点的，会左右摇摆着下落。

如果从足够高的地方释放纸片，或者释放时让纸片竖着且长边对地，那么下落一定高度后，纸片会翻滚下落，并且往左侧或右侧飘。

释放时，如果让纸片下侧的边缘朝左偏转一点点，就会顺时针翻滚向左飘落；让纸片下侧的边缘朝右偏转一点点，就会逆时针翻滚向右飘落。

原来，这种事先让纸片偏转，或者下落过程中因为扰动或不平衡随机地出现偏转，就会使纸片与下落方向产生迎角。迎角的升力效应使纸片受到垂直于下落方向的力，使纸片侧飘；迎角的抬头效应使纸片翻滚。

翻滚下落本身又代表了一种旋转，引起马格劳斯效应，这种马格劳斯效应对应的侧向力，加剧了侧飘。树叶形状更复杂，除了有长方形纸片那样的情况，还会产生轻易捉摸不透的力，因此下落时还可能出现飞舞现象，即突然快速翻滚。如果遇到什么大气尘卷风，那么树叶会在尘卷风中时而飞舞，时而落地，时而迅速飘起。

曲面上的弹球（图1.19）

也许没有比曲面上的弹球更好的例子来帮助理解什么是稳定状态和

不稳定状态了。

将一颗弹球放在平坦的桌面上，轻轻推一下，弹球移动到一个新的位置，停在新的位置上，不会自动回到原位，也不会无限地远离原位。这种情况叫中性稳定。如果将弹球放在球形锅底这样的凹面的上面，轻轻推一下弹球，弹球在重力作用下最终会回到原位。这种情况就是稳定的。如果把锅翻过来，将弹球放在其凸面上，轻轻推一下弹球，弹球就会在重力作用下滚下去，再也回不到原位。这种情况就是不稳定的。

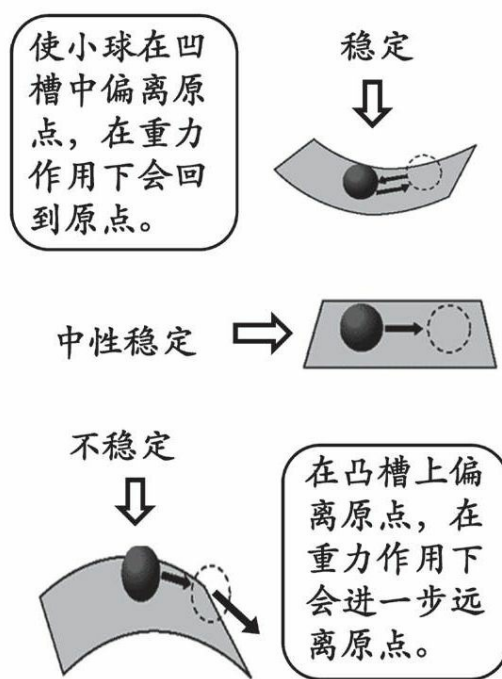


图1.19 曲面上的小球

更复杂的问题中的稳定与失稳，道理类似。失之毫厘谬以千里，指的就是处在一种不稳定状态，稍有偏差就会导致不可复原的变化。

失稳中的多姿多彩（图1.20）

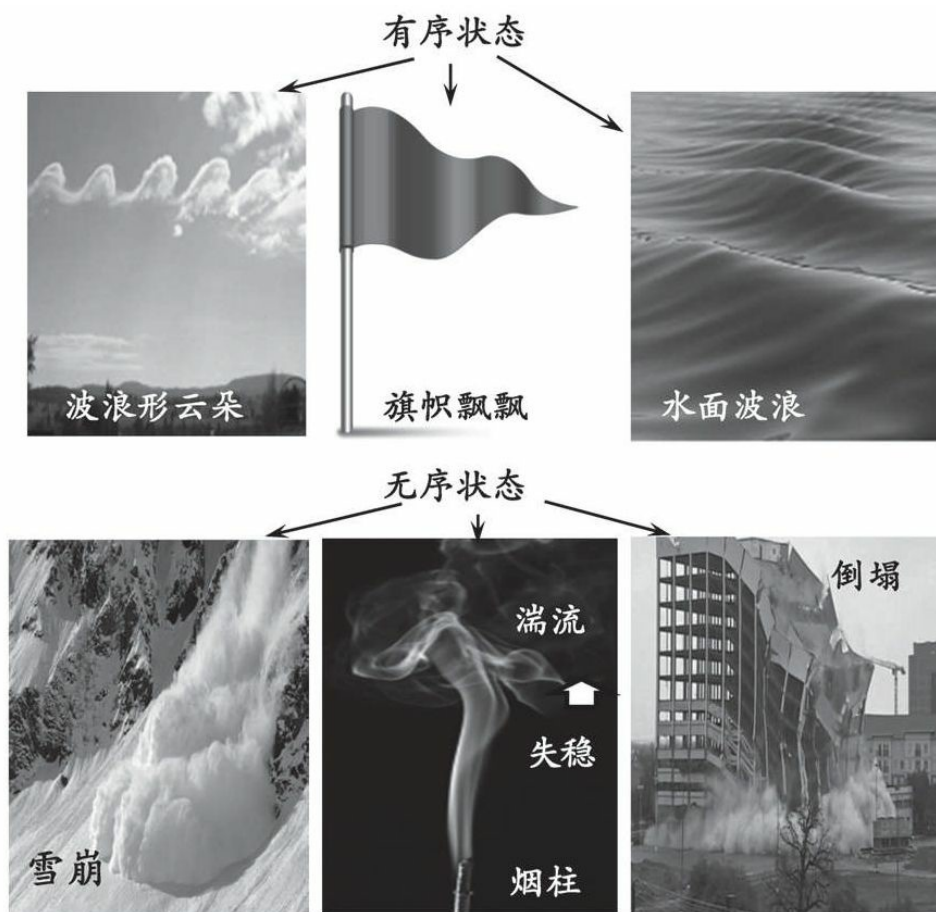


图1.20 失稳导致有序状态或无序状态

自然界也像人，不满足于单调与乏味。美丽的形态，往往是一种更和谐更有序的状态，因此就容易从一个单调乏味的状态通过失去稳定演化而来。最直观的例子就是波浪，如大气波浪云、水面的长条形波浪和旗子飘扬形成的波浪。这些波浪也称为开尔文波，因为开尔文解释了其产生的原因。

看似一串蘑菇型花朵的波浪云，是由单调乏味的两股速度不一样的气流相互搓动引起失稳导致的。

假设高空有一股冷空气从远方吹过来，在低空的热空气上面流过。两股气流相互搓动，总是会引起它们交界面的变形，导致相互渗透。

我渗透一部分到你那边一侧，对你而言就相当于鼓起来一个小鼓包，迫使你走弧线。你就产生离心力了吧，于是你就得降低气压来补偿离心力。你气压小了后，抵抗我的力气就更小了，于是我得寸进尺，进一步深入你的腹地。当然你也不甘示弱，把我突进去的部分吹弯，把我突进去的部分吹得滚起来，让我在那里打转转。把我变成旋涡了。你来我往，我从一个位置插进你的同时，也把边上的你挤得有一部分来不及吹走，来不及吹走的走投无路，只能往我这边挤，突进我的腹地。我也如法炮制，先假装让你，实则也在通过走弧线产生离心力降低气压，把你勾引过来多点，再猛地把你吹翻。

因此，两股有速度差的气流的交界面是不稳定的。一旦出现一侧向另一侧鼓进去的情况，就迫使另一侧走弯路，产生离心力降低气压，使鼓起效应放大。

波浪云也好水面上的波浪也好，为何有时是一种有序的、有浪峰有浪底的一段段谐波（即看上去和谐变化的波）？尤其是，波长不是随意的，而是发展出一个特定的波长（相邻两个浪峰之间的距离）。

原来，初始变形的波长如果太大，那么鼓出来一点点，相对弯曲程度就小，从而离心力不足以驱动产生快速的失稳运动。如果波长太小，就像车行驶在铁轨上不会产生离心力一样，也不容易失稳。因此，一定有一个特定的波长，最容易失稳，或者说失稳后变化最容易快速发展。那些失稳不容易的杂乱变形，变化不怎么快。因此，一定是那种本身就像波浪一样的且具有合适的波长的变形得以快速发展，最终形状被它决定。

当然，气流交界面的初始变形不会恰到好处地是拱桥路面那样的弧形表面。但任何变形可以看成许许多多的波浪形变形叠加而来。这种从

随机的初始变形中“选择”其中包含的令失稳发展最快的波浪，也可以理解为一种优胜劣汰。

在波浪云中，热空气遇到冷空气后，其中的水蒸气凝结成白色水雾，随波浪一起运动，因此就形成了看得见的波浪云。

名片的翻滚下落以及烧开水形成六角形对流腔等，也是失稳引起的多姿多彩的有序现象的例子。

蝴蝶效应 失稳导致的险象与混乱

一只蝴蝶在巴西扇动翅膀，就有可能在美国的得克萨斯引起一场龙卷风。这就是所谓的蝴蝶效应。当然这只是说有这种可能性，实际上发生的概率极小。蝴蝶效应是指某些不稳定系统，稍微受到点扰动，就会失稳导致变化被放大的现象与混乱。

其实，蝴蝶翅膀扇动很慢，每秒也就5~10次。蝴蝶那么小，扇动翅膀时引起的气流与气压变化当然很小，怎么可能诱发龙卷风呢？

在平路上要推动一辆大卡车当然很困难。如果这两大卡车悬在悬崖边上，你轻轻一推，就掉下去了，因为卡车所处的状态是极不稳定的。如果卡车停在平坦路面上，你轻轻一推，它纹丝不动，因为卡车所处的状态是稳定的。同理，如果大气条件恰好处在一个极度不稳定状态，蝴蝶效应就有可能发生。

雪崩、烟柱的上层变成湍流、房屋倒塌、骚乱、踩踏事故、连环车祸、坠机事故、波涛汹涌等，均是失稳导致的险象或混乱。

用平底锅烧水，水慢慢加热，我们从上面可以看到被加热的空气蒸发形成的白雾，水可能在慢慢流动，但幅度不大。如果持续一段时间，水会突然沸腾起来。这表明失稳有可能特别快。

让不稳定变得稳定（图1.21）

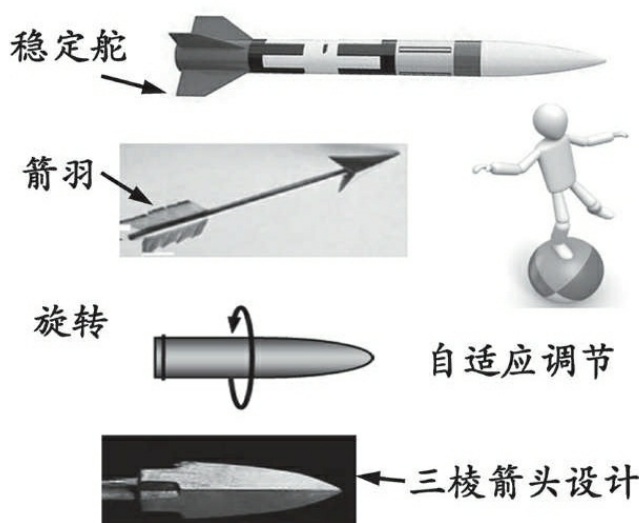


图1.21 让不稳定物体变稳定

有迎角的机翼产生抬头力矩，因此飞行的机翼一般是不稳定的。为了稳定，就需要采用控制措施。水平尾翼包括鸟的尾翼就是用于控制这种不稳定的。机翼抬头的同时平尾也抬头，机尾的平尾增加的升力相对于飞机重心的扭矩就与前方机翼增加的升力相对于飞机重心的扭矩大小相等，方向相反，相互抵消，保证飞机稳定。

不难理解，导弹、船舶、火箭和弓箭，也是通过稳定舵或箭羽来实现稳定。稍微偏离方向，稳定舵就获得迎角，产生恢复姿态的扭矩。

秦弩采用三棱箭头，也能帮助增加稳定性。道理稍微复杂点。标枪

则因其使用性质，不能加稳定舵。通过改变质量分布，让质心靠前，这样遇到偏离方向的扰动，产生的迎角升力近似作用在质心上，就不会产生失稳的扭矩。

某些子弹在出膛前令其高速旋转，这样不容易跑偏，原来，旋转也可以帮助物体运动时更稳定。

2. 旋转助稳 令人痴迷的玩具陀螺

玩具陀螺因为自旋，立在地面而不倒。它历史悠久，让小孩和大人着迷，尤其让科学家痴迷。类似地，飞碟因为自旋，能稳定飞行。水漂如果自旋，不容易跑偏。旋转似乎携带了某种力量，当受到外部干扰时，这种力量就出来帮忙，通过帮助物体各部分绕弯子来躲避、化解、转移或吸收外力在一个点的死缠烂打。在气流中滚转飞行的物体，如旋转的足球，还会具有前面所说的马格劳斯效应，让物体能躲开正面的拦路虎。看来旋转的力量不可小觑。

惯性和转移性让旋转物体更稳定（图1.22）

速度是一种惯性。物体有了速度，如果不通过外力作用，速度大小和方向就不会变化。同理，旋转也是一种惯性，是旋转物体各点速度惯性的一种综合效应。一个旋转的物体如果没有外力干预，就会保持原有姿态继续旋转，即具有旋转保持性。

不但旋转速度即旋转的快慢是惯性，旋转轴即旋转方向也是惯性。没有外力作用，旋转速度和方向就不会改变。反过来，如果临时遇到一点外力破坏，就会牺牲一点旋转惯性，把外力的影响吞掉了。不会进一

步使姿态失稳，因为进一步改变旋转的姿态，还需要新的外力作用。

不仅如此，旋转还可以转移力的作用。如果有持续发生作用的外力试图使物体偏离姿态，不旋转的物体就在一个固定位置或区域持续感受外力的作用。如果属于那种穷追不舍的外力，你越偏离原有的姿态，这个力越起劲的话，那么物体可能失去平衡，甚至被掀翻。

反过来，如果物体足够快地旋转，那么物体的不同部分就会轮番过来感受这个力，而不会停留在原始状态让外力往一个点和一个方向死缠烂打。

石片飞行时，稍有偏转就产生迎角，就会产生升力效应和抬头效应，使石片可能失稳和翻滚。如果高速旋转，抬起的一侧很快转到侧面或后面，不会在原有方位持续接受抬头效应，很难被掀翻。打水漂让石片旋转，就会让石片飞得更稳定。自旋的飞碟更能稳定飞行，也是类似道理。

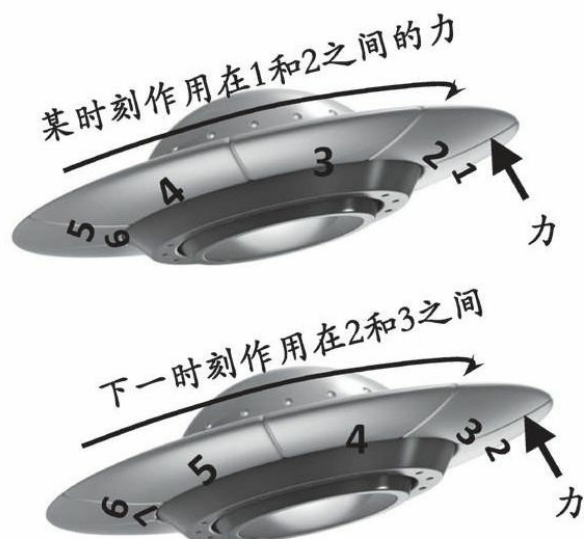


图1.22 旋转的转移效应

不倒的旋转陀螺（图1.23）

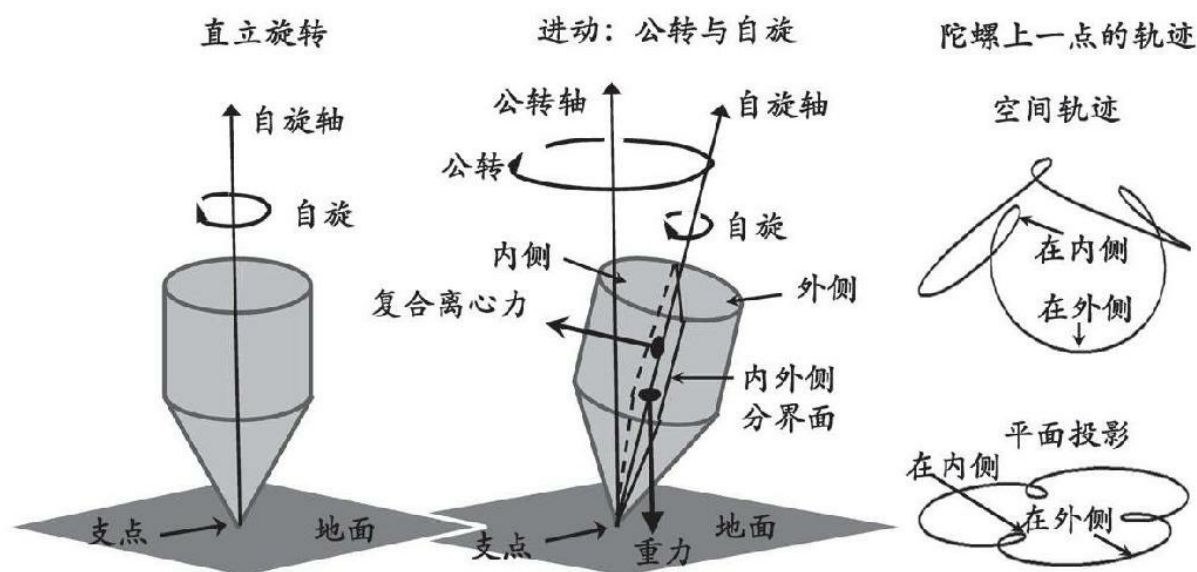


图1.23 旋转的陀螺与进动

玩具陀螺直立情况下，重力正好落在支点上。如果因为某种干扰或者重心不完全在中心轴线上而导致轻微偏转，那么重力就偏离支点，重力就提供了扭矩，进一步扳倒陀螺。

陀螺如果高速自旋，绕初始自旋轴的旋转惯性不会无故消失。如果扰动使自旋轴偏转，陀螺就产生一个绕原有自旋轴（竖轴）的整体旋转（即公转）来尽量保留原有的旋转惯性。于是，稍微倾倒的陀螺，不仅在绕偏斜了的自旋轴自旋，而且自旋轴也在绕竖直轴（公转轴）旋转，这种现象称为进动。自旋速度越快，进动转速越小。从上往下看，公转（进动）的旋转方向与偏斜之前的自旋方向一样，要么都是顺时针，要么都是逆时针。自旋轴和进动轴的夹角有时也会发生变化，这种变化称为章动。

如果自旋速度足够快，那么产生进动后，进动转速与自旋转速成反比，这样才能平衡。如果遇到较大扰动，迫使陀螺轻微倒向一侧，那么重力会试图将陀螺进一步扳倒。陀螺在高速自旋时，重力居然不能把它扳倒，好像重力消失了一样。你往倾倒的方向扳倒它，陀螺就随进动轴转到另一方位，好像旋转的陀螺绕着进动轴在躲猫猫，重力在哪里，就往边上躲。这令人感到惊讶，很难理解进动是如何化解重力影响的。

一种可能的原因（图1.23）

在陀螺某位置上刻下一个记号。这个被标记号的点称为标记点（也称为物质点）。自旋的陀螺在进动时，标记点不仅在绕自旋轴转，自旋轴也在转。于是，标记点不是在做圆周运动，而是在走一条弯曲程度随标记点的位置变化而变化的曲线路径。不难证明，当标记点在远离公转轴那一侧即外侧时，路径弯曲程度小（即相当于绕半径较大的圆），因而受到的指向外侧的离心力小。

反过来，如果标记点在靠近公转轴那一侧即内侧，那么路径弯曲程度大（相当于在绕半径很小的圆），因而受到的指向内侧的离心力大。

在同一时刻，不同标记点（即陀螺上所有的点）有的在外侧，有的在在内侧，离心力大小不一样。靠近内侧的那些点，由于运动路径的弯曲程度大，因此离心力大一些。于是，不同标记点的离心力合起来构成的总的离心力，即复合离心力，不会等于零，会指向公转轴，正好平衡重力的力矩。这就是进动中的陀螺不会被重力力矩绊倒的最可理解的原因。

物理学上的标准解释比这抽象得多。这个抽象解释利用所谓的角动

量守恒原理，角动量守恒原理是牛顿定律用到旋转物体上的一种推广。

永恒的科学与应用价值

陀螺效应既神奇，又难以琢磨，因此吸引了一些伟大科学家研究。例如，1788年拉格朗日就研究过对称自旋陀螺问题。量子力学和原子理论的奠基人，慕尼黑学派掌舵人，德国理论物理学家索末菲（Arnold Sommerfeld），与栽培过他的克莱因（Felix Klein）在1897—1898年间发表了涉及陀螺数学理论的两部专著。有趣的是，索末菲带出的弟子有五名诺贝尔奖获得者，他本人虽然获得81次提名，但从未获得过诺贝尔奖。这么大牌的科学家都研究过陀螺，说明问题多么难多么有趣。

由于陀螺问题非常巧妙，涉及极其有趣的物理学原理，有关陀螺研究的论文成百上千。虽然物理学上能完美地解释陀螺进动现象，但人们不满足于此，总是想找到更直观的方式来理解。国外一些物理学讨论网站，有关这样的问题与回答没完没了。我们总是很难用直觉去理解。也许上面用复合离心力不为零（来自于不同物质点在同一时刻路径弯曲程度不一样从而离心力不一样）的解释属于最直观的一种。

玩具陀螺这样的通过进动来抗击外力干扰的神奇稳定效应，在生活与工业中有大量应用。陀螺仪就是一种很普遍的应用。

让飞行的石片或水漂在出手一刻旋转起来，也存在这种旋转稳定效应，能更平稳地飞行。《神勇投弹手》那部影片中，片中主角“傻子”向敌人阵地甩手榴弹，飞出去的手榴弹在高速旋转，这是为了让手榴弹更稳定地前飞。一些炮弹出膛时，也高速旋转，利用陀螺效应来保持飞行方向。有的子弹在出膛时绕中心轴线高速自旋，就是让子弹飞行更稳

定，保持方向，飞得更远。

3. 振动的世界 共振

在圆底锅里滚一颗弹球，弹球沿着锅底弧面，爬上一侧，似乎打盹了片刻，又徐徐朝锅底滑落，到了锅底，凭惯性又急速地滑向另一侧。如此不断反复，就是一种振动。如果没有摩擦消耗惯性，这种振动就会持续下去。类似的振动，包括弹簧的振动、单摆的运动、琴弦的振动、水面的波动和空气气压的振动等，无时无刻地丰富大自然的运动形态和影响我们的生活。人和动物的腿走路或奔跑时，会同时利用单摆运动和弹簧运动，可见表面上看上去与振动无关的腿的行走姿态，可以看成两种振动形式的叠加。

我们熟悉的单摆（图1.24）

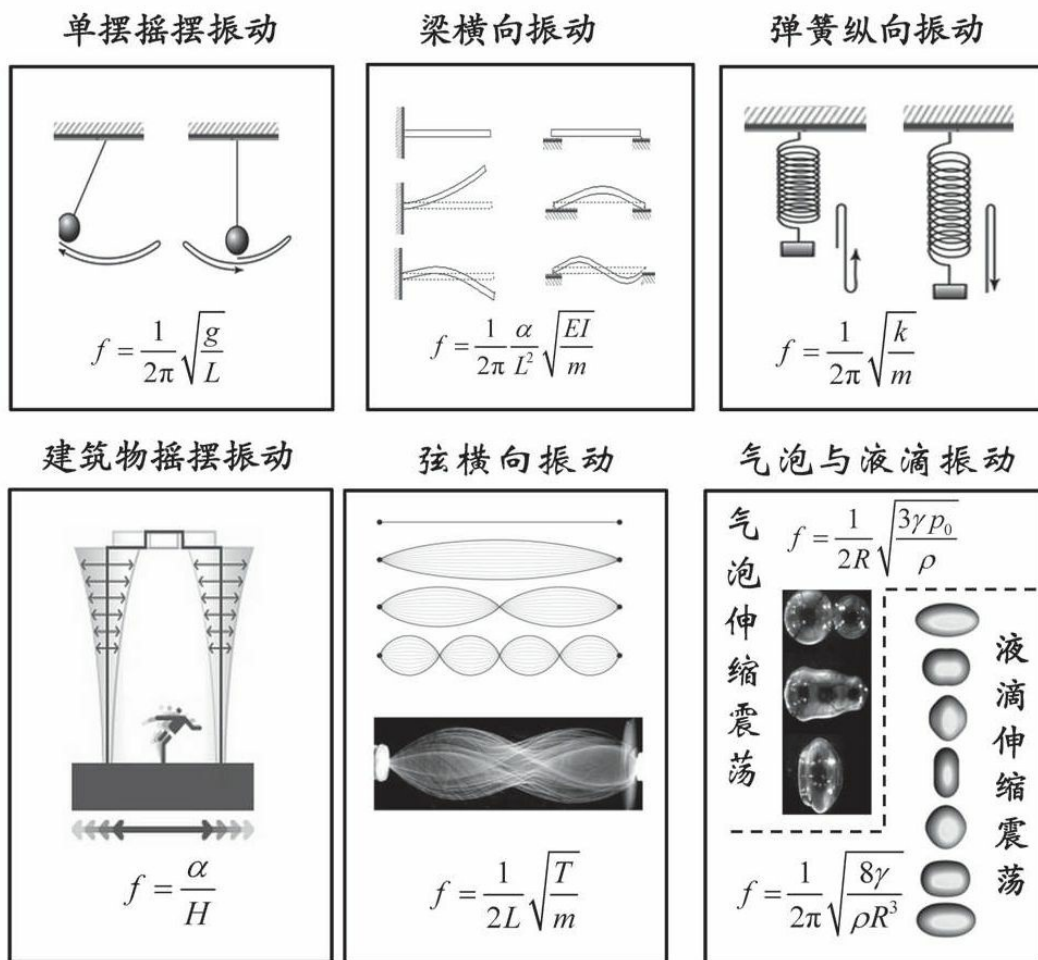


图1.24 振动与震荡的众生相

单摆由摆球和悬挂摆球的摆线组成。摆球受到地球引力即重力的作用。如果把摆球向一侧推高，那么地球引力会迫使摆球随摆线摆回原点，即摆到摆球的最低位置。由于地球引力的加速作用，摆回原点时获得了足够的速度。这种速度蕴含的惯性，会驱动摆球向另一侧抬高。如此周期性地反复，就是单摆运动。

这种来回摆动一次的时间也称为周期。一秒摆动了多少次也称为频率（一般记为 f ），单位是赫兹，1赫兹就是1秒摆动1次，10赫兹就是1秒摆动10次。其他形式的振动也是这种周期性的运动，振动的快慢也用

类似的频率来衡量。

显然，摆线长度越大，摆动走的路就越长，来回摆动一次需要的时间就越长。如果地球引力越大，那么对摆球作用就越大，摆动一次需要的时间就越小。因此，不难理解，摆线长度越大，周期就越长，或者说频率就越小。这种长度或者尺寸越小那么振动频率越大的规律还很普遍的。你可以很简单地得到周期的大致估计。我们知道，物体有重力势能，从一个高度落到另外一个高度，重力势能减小的量等于重力加速度（约为9.8米/秒）乘以掉落的高度，再乘以物体质量。由于能量守恒，重力势能的下降等于动能的增加（质量乘以速度平方的一半就是动能）。如此不难得出速度正比于下落高度的根号（即开平方）。振动过程获得速度，也是通过这种势能与动能的交替转换，因此也与高度变化量（对于单摆问题，正比于单摆摆线长度）的根号成正比。而周期等于高度变化量除以速度，于是周期也正比于高度变化量的根号。作为周期的倒数，频率当然就与高度变化量的根号成反比。用到单摆上，就是与摆线长度的根号成反比。

单摆到了最低位置，获得了速度。由于惯性，驱动单摆向另一侧抬高。下落时速度越来越大，当然上升时速度就越来越小（因为重力朝着向下的方向拉摆球）。单摆最终摆到摆球速度为零的高度。接着在重力作用下，又重复这样一个过程。将物体举高了后，在地球引力作用下要向低处运动，并且到达的高度越低速度越大，这种现象可表示为重力势能的释放。用能量守恒来概括，说明在更高的高度有种能量，否则到了低的高度，动能从哪里来。这种与高度有关的能量就是势能。在没有约束的情况下，高势能需要向低势能位置运动，这与苹果会掉下来道理一样，也与高温的暖气片向低温的房间散热类似。

万变不离其宗的振动世界（图1.24）

物体的一部分偏向一侧，一种力就把它往回拉。回到原点时却没有其他力将其制动。于是到了原点速度就很大，惯性使其偏向另一侧，继续被同样的作用往回拉。于是就像荡秋千一样振动起来。我们的身体也有振动。心脏和脉搏每分钟跳动75次左右。

弹簧、乐器的弦、大风中的高层建筑、湖面的水位、空气中的气压、小水滴和小气泡的形状也会出现类似的振动或震荡或波动现象。道理与单摆均类似，只是驱动振动的因素不一样，振动方向不一样，尺寸选择不一样（有的振动路径是长度，如单摆；有的是一个体积大小在振动，如液滴；有的是物质点在振动，如弹簧和梁）。

微观世界也有振动现象，比如说阳光是由光子组成的。光子也在振动，称为波动。也许是空间在振动，让处在其中的光子振荡起来。对于这样的微观世界中的粒子，振动也成了一种惯性。没有外部作用，这种振动就会持续。振动也代表了一种能量，能量守恒使振动能维持不变。

显然，物体越小，越容易完成一次振动，因此振动的快慢即频率，与长度或与物体的质量成某种反比关系。于是，越小的物体振动的频率越高。我们能看见的光的振动频率，高达 10^{14} 赫兹。而高层建筑的振动频率，可能只有1赫兹左右。

例如，高楼的最低振动频率与高楼的高度近似成反比，具体大小当然还与建筑结构有关。楼越高，振动频率越小。比如说，戴姆（Clive L. Dym）针对某40层120米高的建筑，得到的最低频率大概是0.4赫兹。如果频率与高度成反比，那么20层的楼就是0.8赫兹，10层的楼就是1.2赫兹，8层的楼是1.5赫兹。

水面波动是水位高度的振动，驱动因素要么是表面张力，要么是重力，或者两个因素兼而有之。

并不复杂的复杂：谐波振动与叠加（图1.25a）

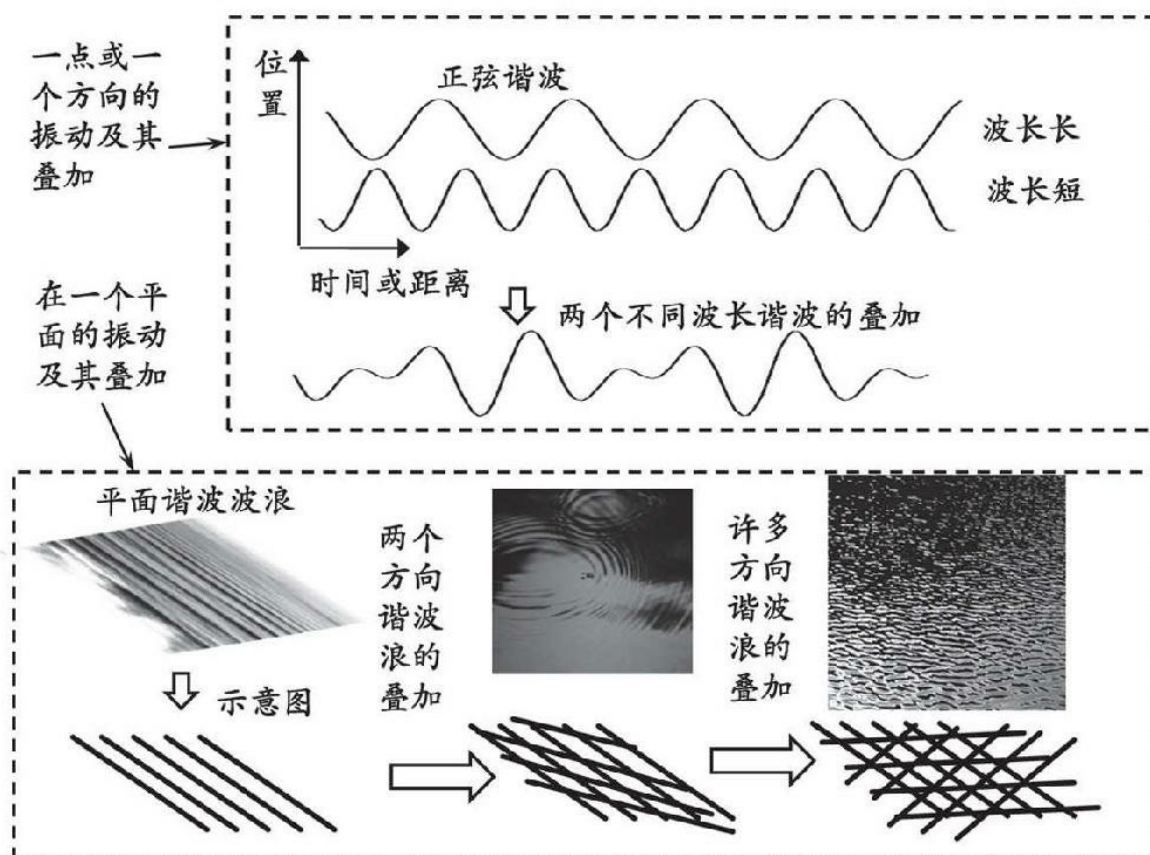


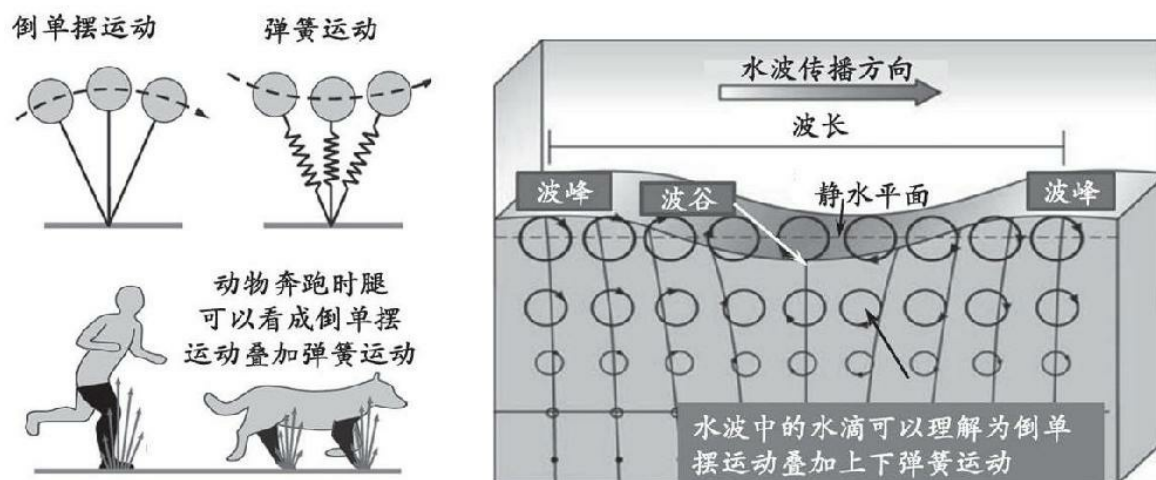
图1.25a 复杂的振动可以看成简谐振动的叠加

弦在做振动时，最简单的形状是正弦曲线。弦上一点的高低位置也是时间的正弦曲线。类似的振动形态称为谐波振动。最高点称为波峰，最低点称为波谷。相邻两个波峰之间的距离称为波长。如果波峰波谷沿着轴线在移动，称为行波，即行走中的波（如水面波浪）。如果波峰波谷在原地切换，称为驻波，即驻留在那里的波（如琴弦振动）。

其他振动形态往往可以看成若干谐波振动相互叠加的结果。以水面波浪叠加为例，如果有各个方向的谐波型波浪纵横交错地叠加在一起，那么水位形态看上去就像猫爪印。这是因为，不同谐波的浪峰和浪谷有错位，叠加后，哪儿高了点、哪儿低了点就打乱了，就可以产生水位高低不一、像鱼鳞一样的涟漪。就像切一块厚度均匀的肉，一刀刀横着切，或者一刀刀竖着切，只要相邻两刀间距差不多，那么刀印看上去就像谐波的浪谷。反过来，如果按刚才横着切的方式切一遍，竖着切的方式切一遍，再以各个角度斜着切的方式切几遍，那么你得到的是一堆碎肉。这个碎肉状是各个方向的刀印叠加后的结果。

一般有某种外部扰动使振动产生，这种扰动可能不规则，不是谐波形态。这样引起的振动可能显得很杂乱。但杂乱振动可以分解为谐波振动的叠加。如果某一个谐波最不稳定，那么它就被放大。大海中看到的平行排列的长条形波浪，就是这样长出的近似为谐波的波浪。

弹簧振动与倒单摆振动：走路与水波（图 1.25b）



来源：迪金森等在《科学》上的报道

来源：艾德·萨尔蒙《海洋科学》讲稿

图1.25b 复杂的振动可看成弹簧与倒单摆的叠加

单摆在左右摆动，弹簧在上下振动。如果单摆的摆线是弹簧呢？那么就是倒单摆运动和弹簧运动的叠加。想想奔跑时，我们的腿，难道不是倒单摆运动和弹簧运动组合在一起？水面波动，我们看到的是波浪形态向一个方向传播，每一处的水位一高一低在那里振荡。其实，分布在高度方向的一线水滴，也是在做倒单摆运动叠加弹簧运动。这样，每一滴水就在画圆或者画椭圆。每一线水都这样运动，挤得边上的水也跟着这样运动，看起来，水面的波纹就向一个方向传播啦。连人的腿都是这样，何况水。水没有腿，但一线水合起来就是一条腿。

共振

一个物体的振动频率，与自身的尺寸、形状与构造有关。因此也称为固有频率。如果用力不断推秋千，使力的节奏能正好与秋千自己的摆动频率一样，那么秋千摆动的幅度就越来越大。这种施加外部作用的频


率与物体自己振动频率吻合时，振动幅度越来越大的现象称为共振。

开车时，打开天窗，将车开到一定速度时（也许是时速60千米以上），就会听到连续发出轰轰的声音。原来，天窗打开后，天窗口前方的顶盖被气流摩擦生成旋涡的频率，与天窗后方的顶盖反射声波的频率出现了共振。轰轰的声音就是共振的结果。这种共鸣现象也称为罗斯特现象，也称为空腔共鸣现象。一些飞机将武器内埋在舱内，投放时，舱就打开了，也存在空腔共鸣现象。共鸣频率如果与结构的振动频率一致，发生结构与气流共振，可能导致结构破坏。一旦出现这种共振，应改变结构设计，避免结构与气流的共鸣现象。

地面的大风也有振动。这种振动的频率不会是一个值，而是在某个频率范围取值。风的振动频率非常小，频率在0.01赫兹和0.1赫兹之间都可能出现。电线、跨桥和高层建筑的固有振动频率可以小到0.1赫兹，因此容易接近风的频率，在风的作用下出现摇摆。因此，电缆、跨桥和高层建筑的设计需要考虑风引起的摇摆，应尽量避免共振频率，或者采用能抗击共振的设计方案。

低层建筑的频率是几赫兹左右，与风的频率错开了，因此不容易出现风导致的共振。然而，地震的频率却在几个赫兹左右。

地震 地震预警（图1.26）



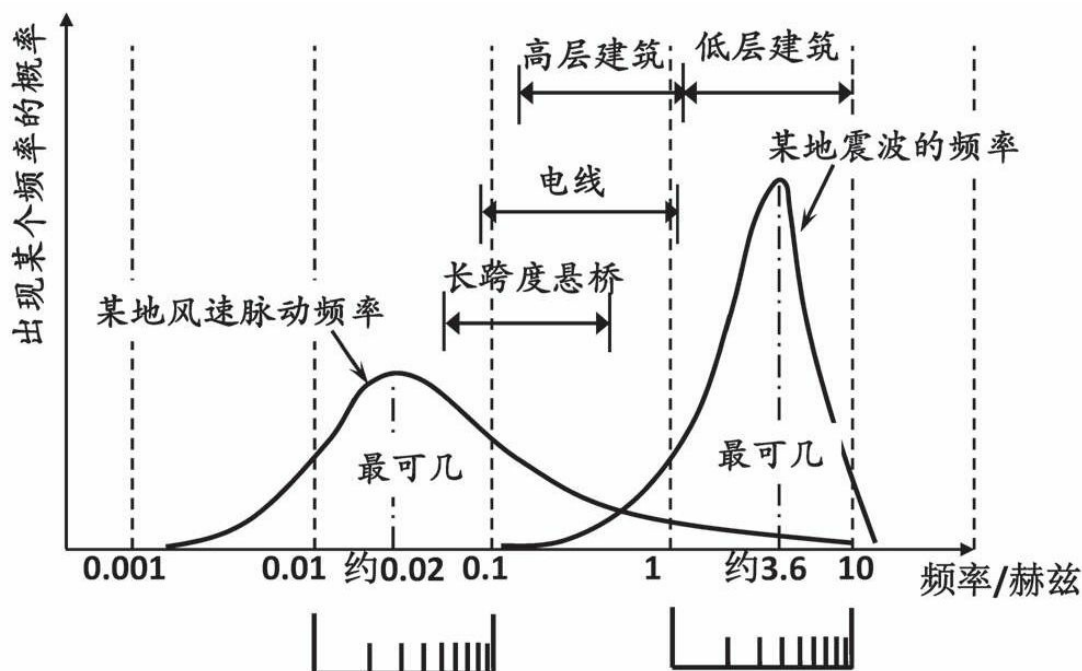


图1.26 不同现象的频率范围可能有重叠

地震是震源中心地壳岩石突然运动引起的地质变形向四周的传播过程。受地震波及的地方会出现剧烈的晃动甚至出现地面塌陷。地震强度用李式级别（由加州理工学院的地震学家查理斯·李于1935年引入）来衡量。加州理工一度将李氏级别、喷气推进以及卡门涡街（柱状物体在风中引起的两串旋涡）作为其主要科学贡献。

地震波分为跑得更快且破坏力不是很大的挤压波（压力波，也称为纵波，因为振动方向沿着传播方向）和破坏力大却跑得稍慢的剪切波（也称为横波，因为振动方向与传播方向垂直）。剪切波是一种频率为几赫兹左右的波，容易与低层建筑物的结构震动频率一致，引起共振，造成较大破坏。

地震导致的剪切波的频率不是单一值，而是在一个范围内均有可能，某个频率出现的可能性最大（称为最可几频率，大致在1赫兹到10

赫兹之间），具体多大还与不同的地震有关。高层建筑自振频率为0.1～1赫兹，低层建筑为1～10赫兹。因此，低层建筑更有可能与地震的频率有重叠，引起共振破坏的可能性最大。据维基百科上地震预警的条目说，八层楼高的建筑物的最低振动频率比较接近地震剪切波的频率，因此更容易受地震破坏。

由于破坏力较小的压力波比破坏力更大的剪切波跑得快，因此异地能提前感知破坏力较小的压力波，如果能提前感受到这种压力波，就可以紧急逃生，在剪切波达到之前，逃生到安全地点。这就是一种预警方式。

也可以在振源探测到地震后，用传播速度为每秒30万千米的电磁波告知远处的人，让远处的人可以提前数秒甚至更多时间逃生，这是更常见的地震预警方式。

2016年10月，日本科学家发表了可以利用高空电离层异常来提前至少20分钟发出逃生的地震预警。他们通过研究发现，至少对李氏七级以上的地震，地震发生前20分钟，高空电离层会出现异常，这些异常可以被立即探测到，从而提前发出逃生预警。

4. 极限飞盘 自行车

说了那么多失稳、旋转与振动，无非是讲了现象与道理。该看看具体的应用轻松一下了，要知道旋转的宇宙还在后面呢。这里就给两个与旋转有关的例子吧，再多，连这里也会不轻松。第一个是极限飞盘。之所以要提它，是因为扔极限飞盘不仅是最近几十年兴起的体育运动，而且简单的形状包含了多种飞行奥秘：既是产生升力的翅膀，也是旋转的飞碟，还可以是徐徐下落的降落伞。第二个例子是与我们生活息息相关

的自行车。让初学者感觉在玩杂技的自行车，其稳定性到底来源于车轮旋转还是什么？科学家为此吵得一塌糊涂。看看他们是如何吵的吧，连超市拖轮车都搬出来了。如果不明白道理，人们就搬出一个更难明白的道理来解释，直到你觉得累了，也就懒得刨根问底了。事情还没完，别忘了你是在空气中穿梭，空气也会添点乱子，怪不得有的自行车会罩一个罩子，说：俺也是流线型啦，空气友好善待我吧。

极限飞盘（图1.27）

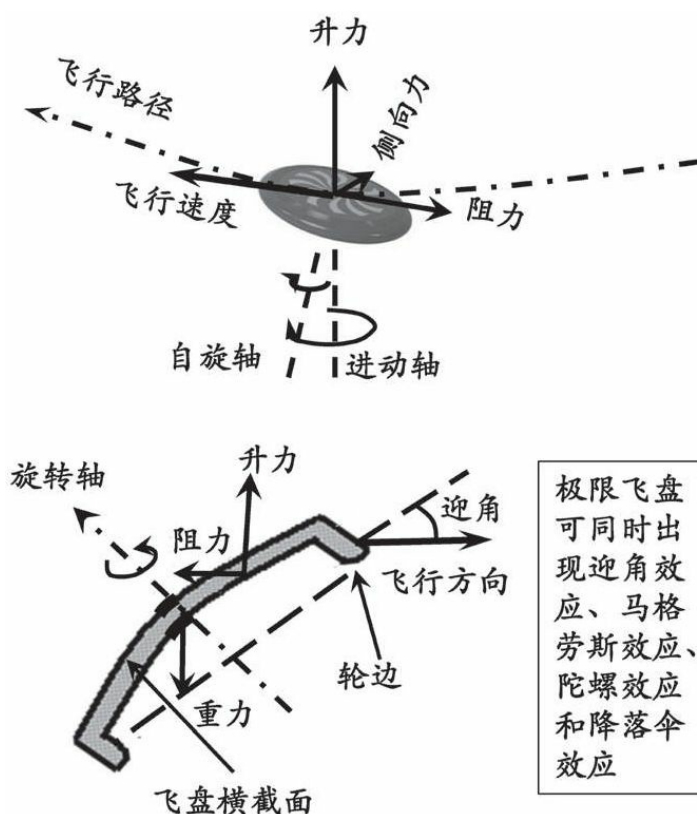


图1.27 极限飞盘

从侧边看飞盘，一面向上凸，近似呈圆弧状，另一面是凹腔。侧边带有轮边，即唇缘。手指握住轮边，甩出去时容易让飞盘旋转起来。从

凸的一面看，飞盘呈碟状，即具有圆盘的形状。直径一般在20~25厘米。由于一侧是带有轮边的凹腔，因此形状有点类似于降落伞。

飞盘形状看似简单，但飞行时包含了极其丰富的科学原理。极限飞盘的飞行，把机翼迎角效应、转盘陀螺效应、旋转时带来的马格劳斯效应以及下落时的降落伞效应集中在一起。

迎角效应：让极限飞盘带迎角前飞，就会产生升力。这种升力会被上表面的弧度进一步加强。由于有了升力，极限飞盘可以飞得更远。

旋转的陀螺效应：让极限飞盘旋转，就不容易失去稳定。旋转时，绕自旋轴有旋转惯性，将维持自旋轴不变，使飞盘能平稳飞行。如果受到特殊扰动，自旋轴被迫改变，那么会产生进动效应，整体上围绕进动轴转，不会完全失去稳定。

马格劳斯效应：由于飞盘旋转，产生侧向力，即马格劳斯力。因此飞盘不会完全走直线，而是类似于旋转足球走弧线。但由于转盘不是特别厚，因此这种马格劳斯效应不是特别强。

降落伞效应：飞行速度降下来后，与速度平方成正比的升力不够了，飞盘理应急速下掉。但由于朝下方的凹腔类似于降落伞，因此下降速度降低。

自行车虽然能稳定骑行，但稳定原理还是引起 争吵

自行车之所以能成为普通人的交通工具，就是因为其小巧可以随意摆放，且省力又稳定，不会轻易摔到。虽然如此，人们会纳闷自行车在

行驶中为何不轻易偏倒，于是自行车的稳定性吸引了较多关注。据说谷歌发明了无人驾驶自行车，如果它轻微倒向一侧，会自动竖起来。

经过适当训练的普通人，骑上自行车，用脚驱动车轮旋转，同时用手握方向盘调整方向不使偏离，就能稳定向前行驶。自行车、方向盘以及身体的协调运动，是自行车不会倾倒的关键。虽然如此，如果行进的自行车自己不具备一定的稳定性，单靠人的协调动作，很难保证自行车不倾倒。于是，人们提出了各种各样的自行车自稳定的理论。自行车自身稳定性吸引了一些著名科学家的研究。

陀螺效应。前面提到的克莱因和索末菲，以及诺特，将自行车稳定性归因于陀螺效应。自行车行驶时，每个车轮绕着自己的轴旋转。因此，如果向一侧微微倾倒，那么就试图改变了旋转轴，就会产生进动。进动试图使车轮自旋轴不会进一步倾斜，而是绕着水平的进动轴旋转，这就迫使微微倾斜的自行车又会竖立起来。至少克莱因、索末菲以及诺特是这么认为的。

离心力效应。著名力学家铁木辛柯和杨提出了一个更好理解的稳定性原理。他们认为，还是骑车者的及时反应与动作起到了稳定作用。如果他意识到自行车会倾倒，比如说向右倾倒，就可以向右打轮，使自行车向右走弧线，这样身体和车产生向左的离心力，将人和车扳正。其实许多人在骑自行车时，可以临时把手离开方向盘，自行车照样稳定地向前行驶。






拖拽效应。琼斯则提出，拖拽效应也使车轮更稳定。如果仔细看自行车，你会发现，挂方向盘的轴是斜的，与前轮的连接点在方向盘的前面一点。这相当于是方向盘轴拖着轮子走。琼斯认为这样产生了拖拽效应。这种拖拽效应起到了稳定前轮的作用。超市的购物车或者货物车，

轮子跟在挂架的后面，实际上是被车拖着走。拖着走的轮子不会走歪，这就是琼斯所指的拖拽效应。

质心平衡效应。荷兰代尔夫特理工大学、特温特大学以及美国康奈尔大学的五名科学家，设计了一种没有陀螺效应也没有拖拽效应的自行车。他们让前轮上下有两个旋转方向相反的轮子，于是陀螺效应相互抵消。后轮也有上下两个轮子，旋转方向相反，使陀螺效应相互抵消。另外，他们让方向盘的旋转轴改变倾斜方式，让前轮触地位置前移一些，就消除了拖拽效应。他们发现，没有了陀螺效应和拖拽效应的自行车，照样可以稳定，前提是自行车质量分布合理，尤其是质心位置恰当。质心位置对稳定性的作用，比陀螺效应和拖拽效应更重要。车前部分的质心位置以及方向盘旋转轴的倾斜方式等，以某种复杂相互作用的形式使自行车稳定。这一结果发表在2011年4月《科学》杂志第332卷上。

显然，不会是上面所说的单一因素起作用。速度的惯性、人的自适应动作协调和上面介绍的几种自稳定因素，在起综合作用，使自行车能稳定行驶。但依据这些稳定性分析建议，可以改进设计，使自行车性能更好。

赛车手分秒必争与自行车的阻力（图1.28）

骑车方式		总重	风阻系数	迎风面积	均速	极速
直身		90千克	1.1	0.51平方米	18千米/时	44千米/时
俯身		81千克	0.88	0.36平方米	24千米/时	55千米/时
双人		163千克	1.0	0.48平方米	25千米/时	59千米/时
倒卧		90千克	0.10	0.44平方米	40千米/时	102千米/时
外罩			0.05	0.13平方米	94千米/时	203千米/时

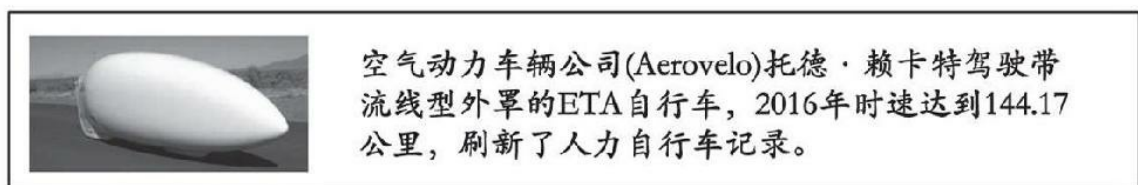


图1.28 让自行车减小阻力提高速度的技巧

另一个问题是，能否减少骑车过程的阻力，可以让我们消耗更少的体力，或者让赛车手取得更好的成绩。

自行车在行进过程中，除了地面摩擦力作用和侧歪时重力作用，还有风阻。自行车与人体作为整体去感受风阻，与石头飞行产生阻力的道理一样。石头飞行受到的阻力与石头形状、迎风面积以及表面光洁度有关。非流线型、表面粗糙以及迎风面积大都会导致阻力过大。自行车也是如此，风阻与形状、迎风面积以及表面材质有关。

为了减小迎风面积，赛车手会俯下身軀，其行车速度至少比常人直身骑车快一倍。如果套一个流线型外罩（如细长纺锤体），那么风阻会大幅度降低。

一般情况下，风阻可占据骑车总阻力的65%~85%，风阻以外的其他阻力是地面摩擦阻力。自行车本身的风阻占总风阻25%左右（进一步

分，车轮占了5%~9%，支架4%~9%，罩6%~9%，其他2%~4%，合起来25%左右），而人体风阻占其余75%左右。

通过俯身可以减小人体的迎风面积，从而减小阻力。车体阻力受表面特性、迎风面积以及形状影响。虽然车体只占风阻的25%，但对于分秒必争的赛车手，减小车体风阻也是必要的。

进一步，如果在大风中骑车，尤其有侧风的情况，那么骑车就困难。侧风产生侧向风阻，容易使车侧翻。如果侧风方向与人的正面有一个夹角，那么就有了迎角效应，有可能产生较大的与侧风方向垂直的力。为了减小这种侧风作用，可以适当扭转身体，使侧风方向与身体的正面尽量平行，这样就可以减小风阻作用。了解这些后人人都是自行车高手，都能像杂技选手一样面对各种情况。

5. 旋转的天体与优美的宇宙

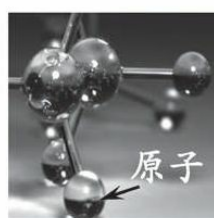
宇宙很大，天文学家有说不完的故事。好多故事我们很难听懂，既深奥又神秘。虽然如此，也有失稳、旋转和振动。按当前学说，宇宙来源于大爆炸，是一种发生在极小的时间尺度和极小的空间尺度上诡异般失稳的结果。数不清的恒星和行星，既不孤零零也不会撞在一起，因为万有引力让已经抱团的星星相互之间不会无限远离，旋转产生的离心力却又避免它们撞在一起。失稳让银河系用螺旋臂秀肌肉。公转让星星构成美丽的星系结构，地球就知道围绕太阳公转，以便不即不离地吸收太阳的阳光从而让大地拥有生命，在太阳保护下不会成为宇宙中的流浪汉。自转让地球拥有白天和黑夜，形成让大自然生生不息的气候。恒星用热核反应照亮了宇宙，它们也有生命的终点。当再也没有热核反应支撑它们庞大的身体后，万有引力会让它们失去稳定轰然坍塌，小的变成

红矮星、中的变成中子星、大的变为黑洞。黑洞非对称地吞并天体会高速自旋，脉冲星的高速自旋让我们能周期性地接受其两极发出的电磁信号，其电磁信号的脉冲性俨然预示脉冲星在振动一样。宇宙天体再大，也是由小得我们看不见的分子原子构成。还有比原子更小的微观粒子与电子，也在振动与自旋，甚至有忍者一样的行为（量子行为）。可以说，宇宙相比于我们人体有多大，那么我们人体就相对于微观世界有多大。总之，宇宙包含了丰富多彩的失稳、旋转和振动等现象。我们先得了解一下，宇宙有多大。

浩瀚的宇宙与万物大小（图1.29）



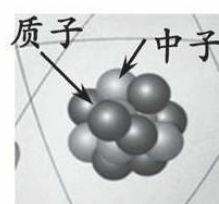
10~20微米
细胞



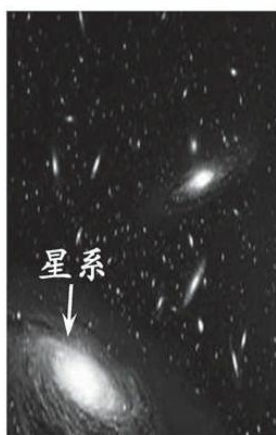
约0.1纳米
分子



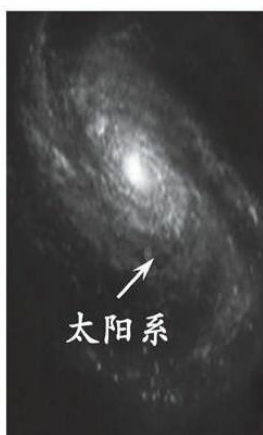
约0.1纳米
原子



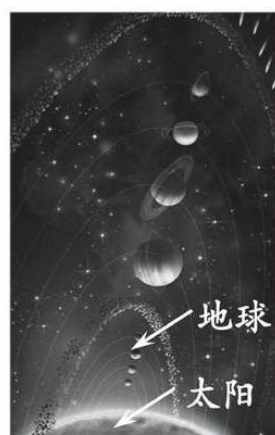
约1飞米
原子核



约数(十)亿光年
宇宙一部分



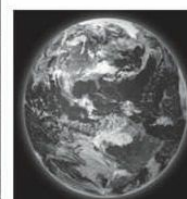
约12万光年
银河系



约100亿千米
太阳系



人 约1米



约1.2万千米
地球

图1.29 万物大小（直径）千差万别

我们人类的尺寸是1米的量级，生活在平均半径6371千米左右的地球上。相比于微米级的微生物，我们硕大无比，相比于宇宙天体，我们十分渺小，哪怕是相对于地球也很渺小。绕地球赤道一圈是4万千米左右。如果你每小时走5千米（据说成人平均速度为每小时5.6千米），且不知疲倦地每日走上8小时，绕地球赤道一周差不多要走三年。当然，千山万水让你走不了那么快，甚至你走不过去。不要这样就觉得地球很大，宇宙比这大多了。描绘宇宙没那么容易，现在连什么是宇宙都有不同看法，更别说有多大。摩尔·萨根的《神秘的宇宙》包含了人类历史。据说宇宙应该包含宇宙遵循的物理规律甚至数学。如果仅仅欣赏安德烈·科尔文（Andrew Z. Colvin）的宇宙图和漫游巡天计划相关的数字化宇宙图（Maps of Universe），那我们看到的就是星系和星星。

我们日常生活看到的物体是当前的物体。宇宙由于太大，大得连每秒近30万千米的光都得走很多年，因此你看到的远方物体，是遥远的过去的物体（是物体发出或反射的光，才让你看得见它）。你现在（用肉眼或望远镜）看到的遥远的天体，是过去的位置和形态，现在可能都到了别处或者消失了。

光一年走的距离，约为94605亿千米（可以简记为接近10万亿千米）。就是说，你现在看到的约为10万亿千米远的一颗星星，实际上是这颗星星一年前的状态和位置。

宇宙从大爆炸开始，过去了约138亿年，这就是宇宙的年龄。可以想象这么长的时间，光线能走多远。138亿乘以光在一年走过的距离，即138亿乘以约10万亿千米，得到的数一口气都念不完。于是，干脆把光一年走的距离称为1光年，于是光线138亿年走的距离称为138亿光

年。这样用时间代表距离，简单多了。

宇宙大爆炸产生的天体物质，按理应跑不过光速。于是，宇宙的半径相比于光在138亿年走的距离，一定小得可怜。

然而，可能不是这样，因为宇宙在膨胀（甚至认为在加速膨胀），星星之间在相互远离。相对于宇宙中的任何一个天体，其他天体都在远离。以我们地球为例，遥远的星球正在远离我们，远离我们的速度与此时的距离成正比（这个观测结果称为哈勃定律，正比关系的比例系数称为哈勃常数）。于是，离地球越远的天体，远离我们的速度越大，而且一定有一个距离，或者一定有一个以地球为球心的球面，那里的星球逃离地球的速度是光速。这个球面称为哈勃球面，其半径称为哈勃半径。在以地球为中心的哈勃球面以外，物质以超过光速离地而去。哈勃半径是140亿光年，比宇宙年龄乘以光速得到的138亿光年距离还要大。

宇宙有多大，目前没有统一的结论。那么，我们能看到的宇宙有多大？即可观测到的宇宙有多大？可观测宇宙是一个以观测者为球心的球体，该球体内所有物体从宇宙膨胀开始到现在所发出的光线和其他信号能在此刻到达观测者，球面以外的则在此刻观测不到。不要想当然地将光速乘以宇宙年龄得到的138亿光年当作可观测宇宙的半径。事实上，考虑宇宙膨胀、光线传播遥远的距离需要时间以及万有引力可以让光线弯曲等因素，可观测宇宙的半径大致为460亿光年。

那由大爆炸产生的宇宙会比可观测宇宙小多了吧。没有定论。有的说小多了，有的说大无数倍（大得几乎没法计数）。到2015年为止，人们探测到的最古老和最遥远的星系（命名为EGS-zs8-1）形成于宇宙大爆炸6.7亿年后，其发射的光用了130亿年到达现在的地球，由于宇宙膨

胀，估计目前在300亿光年外的地方。遥远星系的光线到达地球不容易，因为地球也在随宇宙一起膨胀不断改变位置，遥远星系的光线过来时，还得被引力弯曲方向。

宇宙的大小只能去推测，因为你看不到可观测宇宙以外的东西。那就看看能看得到的天体。可观测宇宙中，有数不清的星星与其他物质。除了我们看到的月亮（地球的卫星）与太阳（恒星之一），以及金星、木星、水星、土星等行星以外，还有数不清的像太阳一样的其他恒星，以及绕恒星公转的行星或流浪汉一样的行星。这些星球不是均匀地分布在宇宙中的，而是聚集成一团团的。

例如，我们地球所在的太阳系就由太阳和几颗绕太阳公转的行星组成。太阳的直径比地球大108倍左右。太阳还不是最大的恒星，有的恒星的直径是太阳直径的1000多倍，体积是太阳的10亿倍（1000的3次方就是10亿）以上。还有许多太阳系这样的恒星系。离太阳系最近的恒星系也在几光年之外，可见恒星之间相距非常遥远。

比恒星系更大的是星系，如我们的银河系，像扁平的圆盘一样（晴朗的夜间遥望天空，许多灿烂的星星在一个条带上，那就是在银河系中的恒星）。星系包含了许多恒星系，星系的直径一般为数万光年（银河系约为12万光年），相邻星系的距离可达3百万光年。在可观测宇宙中，大致有1700亿个类似银河系的星系。小的星系包含数千个恒星，大的可包含百万亿（ 10^{14} ）颗恒星。

许许多多的相邻星系又组成星系群或星系团。星系团直径在千万光年量级，包含数十、数百到数千个星系。包含的星系数目较少时称为星系群。若干相邻星系团还可以组成超星系团。超星系团直径可达数亿光年，包含的星球的质量相当于千万亿个太阳左右的质量。

问题是，这么大的星系，也会旋转吗？

旋转世界的相似性（图1.30）

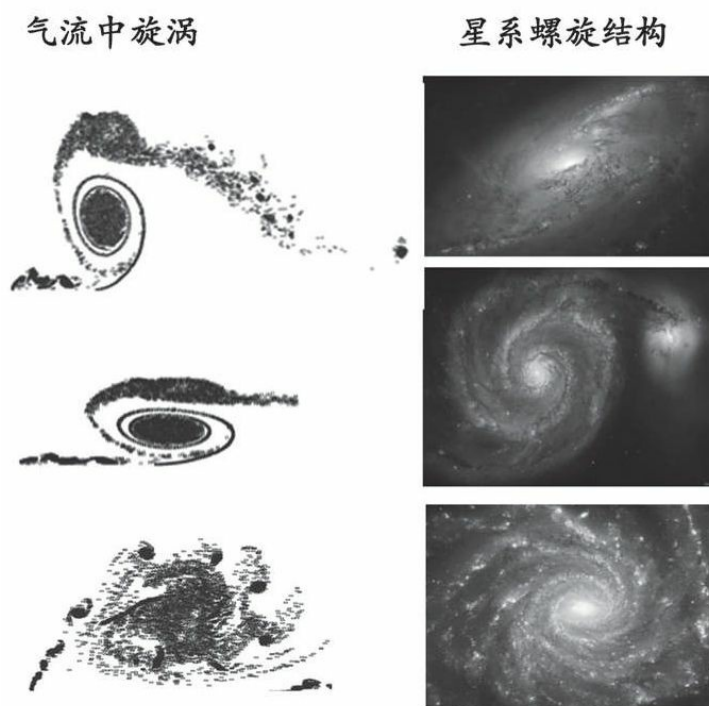


图1.30 旋涡气流与旋转星系很类似

小到翅膀周围的气流，大到浩瀚无比的宇宙，如果我们用某种眼光去观察，会存在一些相似的形态（结构）。运动的形态因此具有相似性，而不管是什么样的物质的运动。例如，让一块带尖缘的平板带大迎角放在流水或气流中，从尖缘会发出一粒粒小旋涡，小旋涡最后会聚合成小的螺旋涡，甚至聚合成大的旋涡结构。这与天文学家观测到的宇宙星系螺旋结构非常类似。自然是公平的，规律不会欺小怕大。宇宙那么大的天体，也存在自旋（如同一粒粒小旋涡）和集体旋转（如同小旋涡随大旋涡旋转）现象。

宇宙中的旋转有三种类型。第一种是大量的构成星系的星球绕着一根中心轴线旋转。第二种是一些星星绕着一颗大质量的星球转，如同地球和火星绕着太阳转，月亮绕着地球转。第三种是星球自转（自旋）。这很像刚才介绍的尖缘发出的旋涡。一粒粒旋涡在自旋（如地球），相邻的小旋涡围绕一个大一点点的旋涡转（如同行星绕着太阳转），大量的旋涡绕着一个中心转（如同星系旋转）。

自转与公转的天体

我们所在的地球就在自转，绕地轴在由西向东旋转，每24小时左右转一圈。我们站在地球赤道的表面不动，被地球自转带动的速度是每秒0.46千米左右。这么快的速度，我们却没感觉。我们不觉得我们在跟着地球自转，反而觉得星星在绕我们转（有星星的夜空，将摄像机对准天空某个方向，每隔1分钟拍一个图像，合起来你就看到星星在绕着地球转）。

高速自转的中子星还会发出脉冲信号，因此也称为脉冲星。脉冲星具有强大的磁场，磁极轴线方向与自转轴方向不一致。这种不一致导致带电粒子与磁场有周期性的相互作用，从而发出与自转周期相关的脉冲信号。脉冲信号极其稳定，比原子钟的稳定度还高万倍以上。脉冲星的脉冲信号的周期有的为毫秒量级，有的为秒的量级。典型脉冲星的半径在10千米左右，质量为1.44~3.2倍太阳质量。发现脉冲星的女研究生贝尔（Jocelyn Bell Burnell）没有被授予诺贝尔奖，而她的导师则因为对脉冲星的贡献获得诺贝尔物理学奖。

如果恒星质量比太阳大许多倍，那么核反应结束从而失去了高温高压的支撑作用后，在巨大的万有引力作用下会坍塌成黑洞。有的黑洞也

在高速旋转。可见，不单是地球这样的行星在自转，一些恒星以及它们坍塌后形成的脉冲星以及黑洞也在自转。

除了自转，天体之间还有公转。地球绕太阳公转，每一年近似公转一周。月球绕地球在公转，绕一周接近一个月。月球一面总是正对着地球，因为月球没有自转。

地球绕太阳公转，离心力平衡了两者之间的引力，使它们的距离不会无限靠近。月亮与地球也是如此。如果没有这些公转，在万有引力作用下，它们会撞在一起。

我们纳闷儿，太阳系为何有公转。其实正是因为有公转，才使它们不会撞在一起。它们形成之时，如果没有公转，要么就撞在了一起形成了更大的星星，要么就足够远不在一个系里了。可见，旋转让我们与太阳不即不离，否则地球不可能有阳光普照，不可能有生命。

巨大无比的星系旋转 暗物质假设（图1.31）

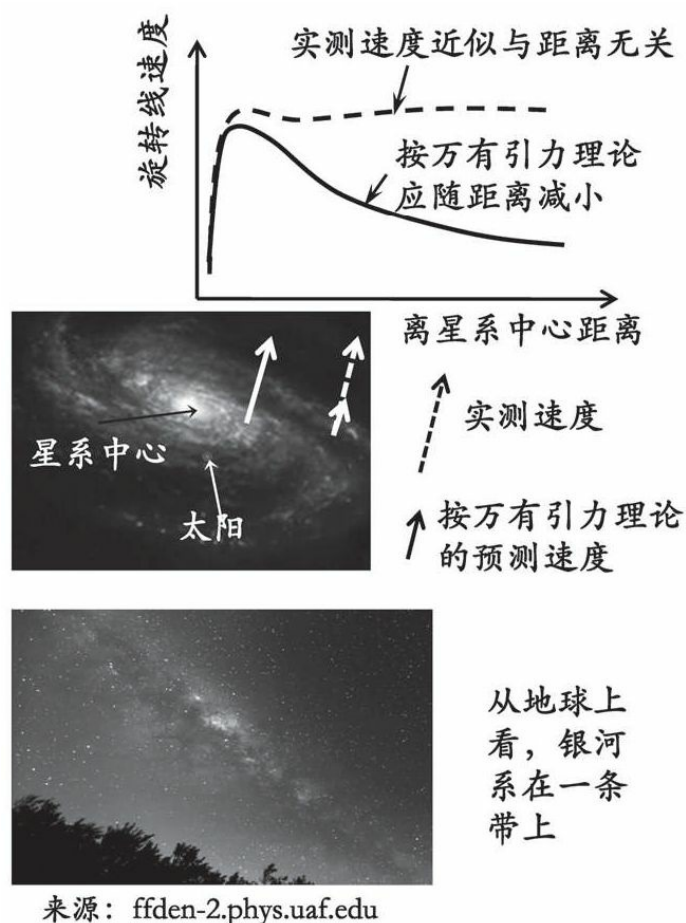


图1.31 螺旋星系及其旋转速度

星系是靠万有引力绑定在一起的一团星球，有银河系这样的螺旋星系（星球绕中心在旋转）、椭圆星系（里面的星球也在按各自的轨道绕中心旋转）以及一些其他不规则形状的星系。星系中大质量天体一般集中在星系中心，万有引力试图将星系拉回到中心，星系旋转产生的离心力则可以抵消万有引力。于是像银河系这样的星系在旋转。

银河系直径达到12万光年，中心厚度为1.2万光年。地球离银河中心（即旋转中心）的距离大约为2.5万光年，地球绕银河中心的旋转线速度达到每秒210~240千米（不同来源给出的这个速度有一些差异），于是，地球绕银河系一圈，需要大约2.5亿年。也就是说，大约2.5亿年

才转一圈。单独看旋转线速度，非常快，但看转一圈多少时间，就显得非常慢。

银河系的旋转会导致一种奇妙的悬臂结构，就像地球大气中的台风结构。螺旋结构的出现也是之前没有螺旋状态失稳后的结果。或者说，带有螺旋结构的星系更稳定，与野渡无人舟自横的道理一样。构成银河系的星球和其他物质从远处看就像一个带有四个旋臂的扁平螺旋结构。银河系螺旋结构大约形成于90亿年前，即大爆炸发生约50亿年后。

星系中心比中心以外的地方亮很多，因此星系的主要质量集中在星系的中心。按万有引力理论（引力与离中心的距离的平方成反比）以及离心力满足的规律（离心力与距离成反比），很容易得出“对于质量主要集中在星系中心的星系，距离星系旋转中心越远的星球，旋转周向线速度越慢”的结论。可是，测量结果表明不是这样，从离开中心某位置开始，旋转周向线速度与距离没有太大关系了。这显然违背物理学原理。

于是，科学家设想，存在所谓的探测不到的不发光的暗物质，它们的额外质量与施加的额外的引力，使旋转速度满足观测到的规律。据此，科学家估算出，宇宙中只有15.5%左右的显物质（即我们可以观测到的物质），其他84.5%左右是暗物质。

也可以从另外一个角度来说明有暗物质存在。由于旋转速度的离心力正好平衡引力，而引力又正比于质量，因此依据星系旋转速度大小可以判定星系中恒星的质量。这样得到的质量称为引力质量。另外，恒星的亮度越大，质量越大，因此依据亮度又可以得出星系恒星的质量，称为光度质量。按理，在排除可能的测量误差后，两个质量应该一样。但实际结果是力学质量比光度质量高了许多倍。因此，科学家推测，力学

质量中绝大部分是看不见的即不发光的物质，即所谓的暗物质。

之所以称为暗物质，就是因为看不见甚至直接探测不到。为了间接探测，就得假定暗物质满足某种理论或模型。例如，目前诞生了许多暗物质理论或模型。其一种是弱相互作用大粒子模型。按照这一理论，这种暗物质粒子会轻易穿过地球而与物体基本不发生作用。这就给暗物质的探测带来了极大难度。我们周围有暗物质流，如果与氙原子发生碰撞，就会发出微弱的闪光和电荷。由于如此的微弱，以致必须在排除了背景辐射和电磁场的地方进行探测。深埋在1.6千米深岩石底下的大型地下氙暗物质探测实验（LUX），力图排除这种背景干扰，通过与氙发生碰撞，探测是否有暗物质碰撞带来的微弱光线。可是，到目前为止，所有努力均未获得任何结果。目前唯一的结论是，暂时未发现弱相互作用大粒子模型是错误的。

暗能量驱动宇宙在加速膨胀吗

去掉所有看得见的显物质和看不见的暗物质后，就是所谓的真空。有理论认为，真空并不真的是空的。真空拥有某种能量脉动，就像水面波动起伏一样。脉动到峰值时，能量转化为一对对正反基本粒子。所谓正物质就是我们能感知的物质，所谓反物质就是能与正物质一起湮灭的物质，如同 $+1$ 遇到 -1 ，加起来就成了 0 一样。能量脉动到极低值时，一对对正反基本粒子又相互湮灭。

人们通过对超新星的观测，发现宇宙在加速膨胀。因此人们设想有种能量，叫暗能量，在驱动宇宙加速膨胀。爱因斯坦说，能量也是一种质量。他导出了一种可将能量折算成质量的关系，称为质能关系。该关系说，能量等于质量乘以光速的平方。如果按照爱因斯坦的质能关系折

算成质量，暗能量估计占据了宇宙中68.3%左右的质量，暗物质占26.8%左右，看得见的显物质只占4.9%左右。

按目前理论，暗能量均匀分布在宇宙之中，虽然密度极低（大约 7×10^{-24} 克 / 立方米），但由于处处存在，总量极大。看得见的每颗星球质量都很大，但它们相隔遥远，因此宇宙中显物质和暗物质的总质量比暗能量代表的总质量小。

可以这样来理解宇宙膨胀。吹气让气球突然长大，气球表面任何两点代表两颗星球，随着气球长大距离越来越大。但这还只是均匀膨胀，加速膨胀就更难想象了。

20世纪90年代奠定的宇宙加速膨胀之说使三位天文学家获得了诺贝尔物理学奖，具体获奖原因是通过观测遥远超新星发现了宇宙的加速膨胀。然而，2016年，尼尔森等三位科学家（J. T. Nielsen, A. Guffanti & S. Sarkar）在《科学报道》（Scientific Report）指出，更多的超新星数据分析提供的证据表明，宇宙膨胀速率是恒定的，而不是在加速。他们认为，过去得到的宇宙在加速膨胀的结论可能来自于某种统计涨落误差。如果是这样，暗能量假设也可能不正确。

微观世界分子、原子与原子核（图1.32）

宇宙那么大，那是用我们人类的眼光去看。可观测宇宙的直径比我们人类的尺寸大了约 10^{27} 倍。宇宙虽然这么大，天体却是由分子和原子等粒子构成。原子是由更小的原子核和电子构成。这些小东西构成的微观世界，与我们肉眼能看到的宏观世界不一样，但也有振动和转动。

我们人体有器官、组织、细胞、DNA。更小的尺度上，是原子。人体大概有1万亿个以上的细胞。据《卫报》报道，人体包含的原子的数目，是 7×10^{27} 的量级。也就是说，一个人体细胞就有数百万亿（约 10^{14} ）个左右的原子。可见原子有多小，尺寸比我们人的尺寸小100亿倍的量级。也许，在分子和原子看来，我们每个人就是一个小宇宙，看上去有星系那么大。还有比原子更小的原子核，平均尺寸比人体尺寸小100万亿倍左右。可以随意穿过任何物质的中微子，其尺寸是 10^{-24} 米的量级。

这还不算最小的，普朗克长度约为 10^{-35} 米。在该尺度上看一个原子核，比可观测宇宙还显得大。在普朗克尺度上看空间，都不光滑了，就像我们的皮肤看上去很光滑，但用显微镜看，我们的皮肤一点也不光滑了。

这种尺度差别太大，还有许许多多的具有不同尺寸的物质，无法在这里一一列举。我们看看分子、原子、原子核和电子就行了。这里不仅有平动，还有振动和转动。

原子由带正电的原子核以及围绕原子核的带负电的电子组成。原子核又由带正电的质子和不带电的中子组成，质子和中子统称为核子。电子绕原子核不断地转动。电子也有自旋。一个质子的直径是 10^{-15} 米左右。电子直径是 10^{-18} 米左右。电子在原子核周围直径约为 10^{-11} 米的球内旋转，这个球就可以看成原子的半径。因此，原子核在一个原子中只占很小的空间，就像空旷的房间中一只苍蝇那么小。因此，像石头这样的固体，表面上看实实的，实际上空得很。星球也空得很，难怪恒星烧完后，会坍塌成尺寸小了许多倍的致密黑洞、中子星或红矮星，这些致密星球，原子核都挤在一起了。

以我们时时刻刻都在呼吸的空气为例。空气分子包括了氧气分子与氮气分子。以一个氧气分子为例，它由两个氧原子组成。

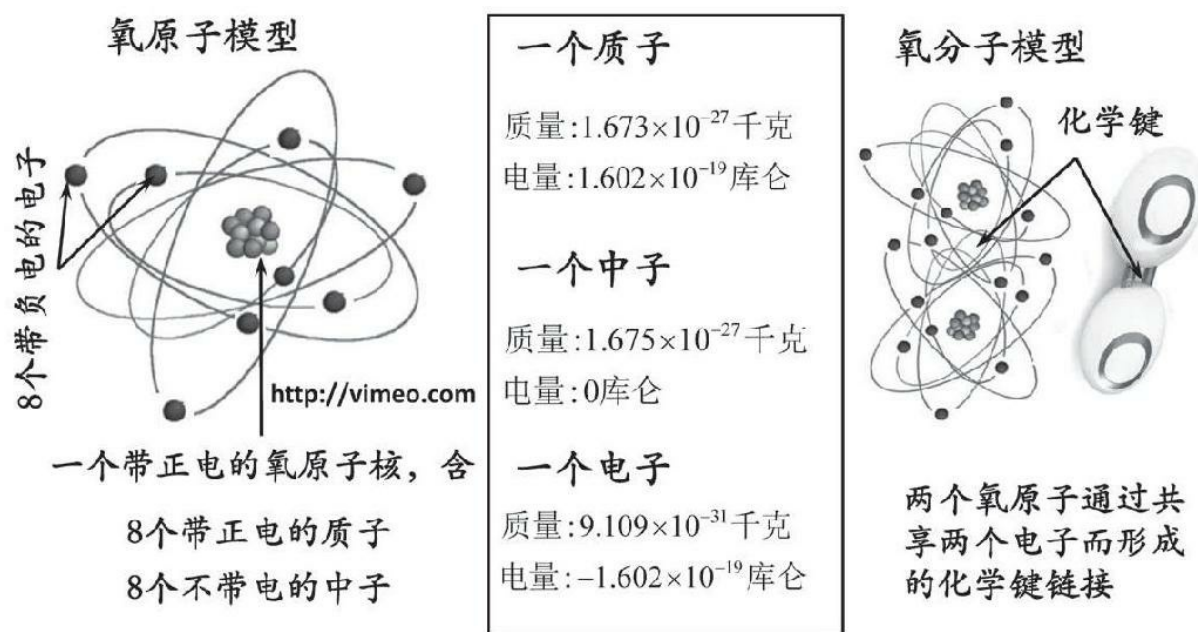


图1.32 氧原子和氧原子结构示意图

别看电子小，它的作用非常大、非常奇特。它携带负电，原子核中有一个质子就有一个电子，你一个质子带多少正电我一个电子就带多少负电，使一个原子不带电，否则我们就会被电着了。电子是红娘，凭借对质子的（静电）吸引力，可以把一个或几个原子胶合在一起形成分子。也可以这样说，两个原子共享原有的几个电子，电子对质子的静电力就把本来会相互排斥的两个原子核死死地拉住了，好像有根很结实的胶带拉住了似的。这根“胶带”也称为化学键。

电子在绕原子核不同的轨道上旋转。如果串到更接近原子核的轨道上，就会发射一个光子。我们的阳光就是由光子组成。光子也有振动甚至自旋。

大气分子吸收阳光中的光子。光子的碰撞让分子拥有平动、转动和振动，这就是分子热运动的来源，是我们周围的大气拥有温度的来源。吸收光子也可能将原子中的一个电子打跑，于是高空大气中就有自由电子以及原子失去电子后形成的离子。光子也可以把一个分子拆开，于是高空大气除了氧分子，还有氧原子。电子如果能吸收光子，那么就会跑到离原子核更远的轨道上，再跑回低轨道时，又发射一个光子。极端情况下，吸收光子还可能引起复合化学反应，即不同原子之间结合成为新的分子。

氧原子的原子核由8个带正电的质子和8个不带电的中子组成。氧原子的电子数目是8个。每个电子与每个质子的电量大小相等、符号相反。因此，氧原子整体而言不带电。两个相同或者不同的原子通过共享电子（化学键）组成分子。共享的电子通过静电力将两侧带正电的原子核吸引在一起，这就是化学键的本质。

一个带化学键的双原子分子，也可以看成一个弹簧振子。空气温度高了以后，每个分子都会像弹簧一样振动。化学键使分子中的原子不容易拆开。拆开就需要能量。化学反应就是分子拆开和合并的过程，会吸收和释放能量。如果是释放能量，就称为燃烧。如果是吸收能量，就称为吸热反应。

一个原子核可能包含了多个质子。按理，带正电的质子之间既有正电相斥的使它们分离的静电力，也有让它们试图聚集在一起的万有引力。可是，质子之间的静电力比万有引力小多了，因此将核子（质子与中子的统称）抱成一团还有别的力，这个力就是所谓的剩余强相互作用力。为何是剩余强相互作用力？原来，质子和中子里面还有小结构，就是夸克。夸克之间由强相互作用力胶在一起。这种强相互作用力还剩下一些，就把质子和中子捆在一起了。

强相互作用是一种短程力，在距离极短时比电磁力大许多，距离较大时就小得几乎不存在了。只有提供足够的能量做功，才能将原子核中的质子和中子拆开。反过来，将孤立的核子结合在一起，会释放巨大的能量。

捉摸不透的微观行为

宇宙万物本身是由原子组成的，如同科特·施塔格（Curt Stager）在《诗意的原子》中所要表达的，自然规律也可以从原子行为诠释。虽然如此，原子之类的微观粒子（除了原子，还有分子以及更小的电子和组成光线的光子）的运动（平动、自旋和振动）状态完全不同于宏观物体的运动状态。微观粒子有所谓的量子行为，其中的自旋就非常有意思。有这种量子行为的粒子称为量子。电子和光线中的光子有这种行为，因此研究这种行为时就把它统称为量子。

量子很诡异，它们的平动速度、自旋速率和位置等遵循量子化行为。比如大气中空气分子的动能，其大小是某个极小数值的整数倍，而不能取其他值。电子绕原子核的轨道也是这样，只能在一些特定的轨道上。吸收一个光子会提升轨道高度。降低一个轨道高度就发射一个光子。除此之外，运动状态和位置还有非常难以理解和描述的不确定性。

以自旋为例，量子围绕自旋轴不是确定的左旋或右旋。微风吹拂下的水面涟漪上的一滴水的位置和姿态在振荡，这种振动可视为由具有振动弦那样形状的谐波叠加而成。如同涟漪中一滴水的这种姿态，量子的自旋是一种左旋和右旋的叠加状态。真不知是微观粒子本身在做这样的运动（我们可以反问驱动它们这样运动的力是什么），还是承载量子的空间就像涟漪，量子只是在那里随波逐浪，以致波浪是怎么叠加而来

的，量子就处于什么运动的叠加状态。你去测量量子的自旋时，就如同测量水波时挡住了一个方向的风一样，由风激起的谐波波浪消失了，于是你就只能测量出一个确定的状态，要么左旋要么右旋，那种叠加状态被破坏掉了，量子力学所称的“坍缩”就成了确定的状态。

水面一个谐波形波浪中的两滴水，a和b，一个处在浪谷一个处在浪峰。如果水滴a随波浪的振荡从浪谷上升到浪峰，水滴b就从浪峰振荡到浪谷，反过来也是如此。于是在波浪中的水滴a和b同步地向相反的方向振荡，如同它们之间有种与距离无关的感应机制。我们可以说它们之间有了纠缠。

与谐波中的两滴水的同步类似，两个量子（量子A和量子B）的自旋方向可以通过相互作用或某种人为方式引起纠缠，即所谓的量子纠缠。相互纠缠的A和B，不管后来被分开多远，如果A被测量时坍缩为一个确定的状态（如左旋），那么B瞬间（没有任何时间差）坍缩为对应的确定状态（如果A坍缩为左旋，那么B坍缩为右旋，反之亦然）。好像有一种远距离感应，量子力学界称为鬼魅般的远距离作用。

当然，以上对量子纠缠的介绍很不严谨，准确表示上面提到的不确定性、纠缠和坍缩，需要使用一些十分难以理解的术语。其实，对量子现象本质的解读依旧争议不断，甚至不同的量子力学专家的解读都不一样。难怪20世纪60年代理查德·费曼说：我敢肯定，没有人真正懂量子力学。量子力学在数学表述上非常完美，可以求出与实际观察相符的解来，但对其中隐含的物理现象的解读如同量子行为不确定一样，不同人理解不一样。连理解都被量子化了。

人们正在探索利用远距离量子纠缠现象进行隔空传输（teleportation）等应用。认为至少可以隔空传输运动状态（而不是能量

和物质)。量子状态是一种叠加状态,利用这种叠加状态可以表示更多的信息,可望出现基于这种特性的量子计算机。量子行为中的坍缩特性导致量子通信很安全,因为一旦被拦截,就像被测量一样携带的信息就坍缩掉了。当然,这些都是正在探索的应用。甚至有人质疑这些探索是否基于对量子行为的某种误解,认为探索不会有结果。

量子行为经常被用于解释一些目前无法解释的现象。一些无法解释的现象经常被臆想是一种量子行为。

世界的大小是绝对的吗

宇宙天体大得惊人,微观粒子小得可怜,那是以我们的眼光去看。也许不同大小的物体感知的时间与空间的方式有差异。也许在光线和电磁波中的光子看来,空间中就没有距离,或者说空间对它们而言,就不是距离。在我们看来不管隔多远,在它们看来也许就在附近,甚至都没有附近的概念。微观粒子之所以在我们看来那么诡异,是因为我们用我们从宏观世界感知的常识和发现的规律去衡量它们。也许因为它们是如此之小,从而感受的空间不是我们感觉的空间。也许都搞不清到底是它们在空间中运动,还是极小尺度的空间本身在运动从而带动它们在运动。

按照之后将要介绍的爱因斯坦的相对论,在以光速传播的光子看来,空间在运动方向变得无穷小了。如果用镜子把光线运动方向偏转到垂直方向,另一方向的空间也变得无穷小了。也许在光子看来,世界应该就没有大小,于是在我们看来的远距离量子纠缠,在量子看来也许还是在一起,甚至就没有在一起和不在一起之说。

单单通过纠缠两个异地的量子除了可能隔空传输状态外，还能隔空传输物质和能量吗？由于这种纠缠发生作用时没有时间差，如果能隔空传输物质和能量，那就意味着物质和能量的运动速度可以超过光速。可是，按照被验证了的爱因斯坦相对论学说，任何物质都不能超过光速。

1.3 运动与演化 过程的激荡与结局的完美

世界千变万化，物体千姿百态。多样的运动用拉长时间去看或者合在一起看就是演化。演化中有速度的增减、数目的涨落、大小的改变、类别的替换和结果的选择。生长与衰减并存于宇宙万物运动及其演化之中，或一枝独秀，或此消彼长，或势均力敌。演化的目的是找到新的归宿，过程可能看似混乱无序，那是为了省时、省力，代价最小化。有时是大自然对简单、简约、简美的追求决定了结果，演化只是去选择获得这种结果的最省时、省力和省料的过程。因此，虽然宇宙由原子构成，它们却不会演化成沙子，而是演化成绚丽优美和多姿多彩的宇宙，包括莲花池长出的是荷花而不是长出污泥。对数目的选择也是一种演化，最后的结局是少了嫌不够多了记不住的数字，数字七则可能是这样一个不多不少的数，于是就有了七色光、七言诗、七音符等。

1. 速度的消失带来损伤 蚊子会被雨滴砸坏吗

获得速度是一种运动形式，让速度突然消失也是一种运动形式。给物体施加力，物体就加速获得速度，这是我们知道的常识。反过来，速度的消失也会产生力。瞬间消失就能造成瞬间打击力。但也不是力量有多大就可以解多大的恨，单手使劲去拍打惹你生气的飞行毛毛和苍蝇，它们就能轻轻地躲开，以柔克刚这个词句大概就这么来的。蚊子比毛毛还难对付，否则怎么在那么频繁的风雨中能生存。小物体碰大物体，速度瞬间消失，给自己带来的伤害更大。这么说来，打大的不行，打小的也不行，打软的也不行，打硬的也不行，难怪人们说君子动口不动手。人体接近2平方米的体表，在承受近两万千克的大气压，平均每平方厘米1千克左右。可见，我们人体内压可以平衡大气压。人能使出的力量

和承受的外力，可以在1000斤量级。千钧之力应该说是这么来的。

雨滴中生存的蚊子（图1.33）

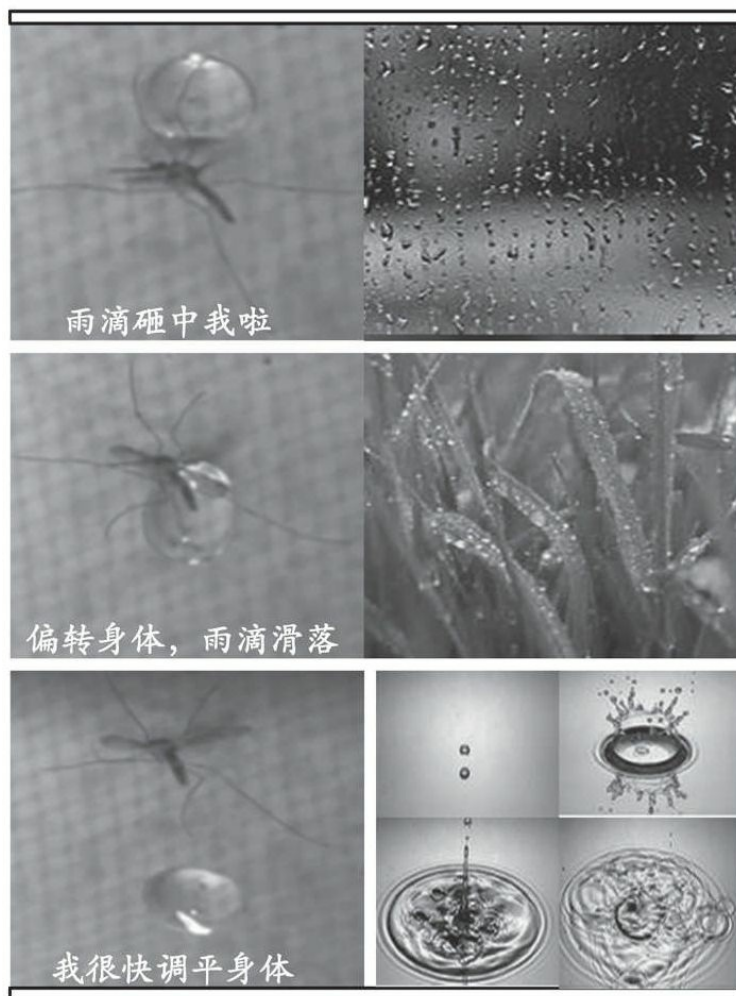


图1.33 蚊子与雨滴

十千米左右高度形成的雨滴，落下来会砸伤人吗？砸不伤体型较大的人，也许会砸扁蚊子。下雨的时候，蚊子是提前躲起来了，还是任受雨滴撞击，以柔克刚地化解雨滴的击打？

胡立德等用实验研究了水滴砸蚊子，得出了蚊子因体重较轻且可以通过偏斜身体来化解雨滴撞击从而能在雨滴的碰撞下得以生存的结论。这项成果让他得到了菠萝科学奖。

雨滴的下落没有想象的那么快。由于存在与速度平方成正比的空气阻力，因此，雨滴落地阶段的终端速度（最后下落的速度）每秒才几米。一般不会超过10米 / 秒。其次，蚊子反应能力极强。我们可能经历过被它侵袭时那种恨不得拍死它却又死活拍不着、关灯入睡时它来了、开灯寻找时不见了的情景。你搞不清它掌握了什么样的飞行奥秘。夜深人静时，它能鬼鬼祟祟地落在你身体上任何部位。等到你有所感觉，已经将毒液注射到你的皮下。

大自然的雨滴总该教训教训它们吧。那么密集的雨滴，就不信砸不中。几十倍蚊子重量的雨滴，消灭不了它们？可是，大多数情况下，只能砸中一侧的翅膀。翅膀结构如同荷叶结构一样，是疏水的，即不沾水的。蚊子巧妙地往被砸中的那一侧歪斜一下，像荷叶上的水滴一样的雨滴，滑走了。这种侧歪应该是被打歪，而不是刻意侧歪来躲避。

蚊子被打歪也没问题，它有不到百分之1秒的反应能力，很快就可以通过娴熟的飞行艺术调整飞行姿态。它的腿也有绒毛，也是疏水的。砸中它的腿，也不会打湿，雨滴滑走了。

即使倒霉地被雨滴砸了个正着，由于它又小又轻，要么被推开（如同你挥拳重重地击打在你眼前飘来飘去的小毛毛），要么被主动侧身躲掉，最糟糕的情况就是被雨滴裹住后一直砸向地面。

如果被雨滴裹住，它正好享受坐降落伞的滋味，但同时也会试图挣脱。随雨滴砸到地面上会发生什么就得看情况了。由于翅膀以及腿的疏

水能力，它不会被打湿，不会被水粘在地上，没被砸晕的话，可能像芙蓉出水一样地钻了出来。

如果被雨滴裹住带入水中，就不知道发生什么了。雨滴落水，会激发涟漪圈圈和碰出小浪花，中间可能有小水珠回弹。被裹住带入水中的蚊子，还能有方向感，还能起死回生吗？

如果还能起死回生，那么可能听到雨滴落水的声音。雨滴落水声音可不那么单调。

有一首“瀑布”（Cascades）歌，里面有一句：让我们倾听落雨的旋律！我们能听到沙沙的落雨声音，啪啪的雨滴拍地声。但是，雨滴落入水中，声音却很奇特。我们人耳能听到水滴落水“咚”的声音，但仔细测量，还有余音。原来，雨滴潜入水中时，会带起一个小空气泡。小空气泡会震动，频率可达15000赫兹，即每秒震动15000次，因此在人的听觉范围内。听到咚的一声后约2万分之1秒内，就有气泡发出的铃响声。

武师打击力

如果水滴撞击地板和人体，也不会带来什么损伤。人是在风雨中进化出来的。如果雨滴斜着击中身体的侧边，则有效撞击速度是投影到人体表面法向的速度。如果雨滴撞击头顶盖，那么由于我们的头盖骨进化得已具有较高抗打击能力，因此也带来不了什么损伤。尤其是击打物体时，雨滴会变形消耗了动能，因此物体受到的力与刚球相比小很多很多。再说了，人体没有想象的脆弱，否则武师就不需要千锤百炼去练成令对手伤筋动骨的绝技。

对普通人而言，武术是神秘的，无论是用于攻击、表演还是养生。

因为神秘所以才导致好奇。将武术赋予了神秘的色彩，才使对手产生畏惧，不战而屈人之兵。泰拳、巴西柔术、少林硬气功、日本忍术、武当剑……。这些名称，本身就冷嗖嗖的。自然，人们关心的是什么武术最厉害，神秘在什么地方。其实，武术的目的不一样，要求不一样。竞技体育中的武术不是以置人于死地为目标，只是体现一种对极限的追求以及体现某种对抗性、表演性和观赏性。非竞技性武术则以修身养性、强身健体和提高自身素养为目的。有的绝技则是为了追求某种人体功能所能达到的最高境界，属于一种形式的极限运动。

科学测量能给出一些可定量描述的答案。《国家地理》“科学看武术”栏目有一组数据，是科学家用科学手段测量几类武林高手的杀手技在面对同样目标时的数据。包括空手道、综合格斗、跆拳道和卡波耶拉的高手都参与了表演和测试。

测试得到的数据包括高手踢腿时的攻击速度、打击力度和击打时的加速度，也包括他们能承受的打击力。武林高手踢腿时，瞬间可以达到时速100英里的量级（1英里约为1.609千米），能产生1000磅（1磅是0.45千克左右，即不到一斤，1000磅是450千克左右）以上的打击力。具体而言，空手道高手西蒙（Simon），踢腿时速为71英里，打击力为429磅。综合格斗高手利维（Levi），踢腿时速为130英里，打击力度为1400磅。跆拳道高手布雷恩（Bren），踢腿时速为136英里，打击力度为2300磅。卡波耶拉高手拉蒂夫（Lateef），踢腿时速为99英里，打击力度为1800磅。

当然，这是针对静止目标产生的瞬间击打力。如果是击打具有躲避能力的手，那么击打力会小多了。这有其科学道理。原来，物体的质量乘以速度代表了一种动量，即动的力量。物体的质量乘以速度平方的一半代表了能量。撞击或击打时，让动量消失需要一定时间的力的作用

（物理上叫冲量定理）。作用时间越短，力度就越大。让动能消失需要力作用一段距离，完成作用的距离越短，力度就越大。于是，用同样的出腿或出拳速度，打在软物体上或者打在可以躲闪或后撤的物体上，作用时间或距离就会更大，形成的打击力量就会小。

人体抗击打能力

人体不像毛毛和小苍蝇那样轻，受到击打时不会迅速跟随打击物体一起运动，而是基本不动地在原地接受冲击力（至少对于普通人是如此）。如果是巨型刚性结构物体撞击地上的没有躲避能力的行人，那么损伤就会极大。汽车如果以每小时100千米的速度误撞行人，那么行人死亡率会达到100%。每小时70千米死亡率达到90%以上；50千米，死亡率为60%；时速40千米死亡率40%；时速30千米为生死临界点。时速20千米，被撞行人也会承受较大伤害。

时速20千米，就是每秒5.5米左右，大致是降落伞落地时最大允许速度，或者人从1.5米高跳下时的落地速度。双脚能承受从1.5米高跳下的冲击力。如果是汽车正面撞击人体，由于人体一般部位接受横向冲击的能力比双脚差，因此被20千米时速的汽车撞击，也会造成较大伤害。

因此，被小物体撞击，或者被变形的水滴撞击，伤害不会太大。但如果是大物体尤其是汽车那样的钢制物体撞击人体，就容易受到巨大伤害。

如果是身体侧面一个点被击打，人体受到击打时会自然侧歪，化解一部分打击力。打击胸部一点也会通过倾倒而化解一部分力。如果击打头部，则由于颈部能传递力的能力较小，因此打击力主要集中在头部发

生作用（包括扭矩作用），容易受到更大伤害。

我们能承受多大的力，与力的作用面积有关。一根细细的针，轻轻用力就能扎进肌肉。青壮年用一根扁担可以挑起100千克左右的担子。扁担压在肩膀上的面积可能就10平方厘米左右，每平方厘米10千克量级。可见我们能承受的外力非常大。我们环境的大气的气压是每平方米1万千克左右。我们人体表面积接近2平方米，因此我们在承受1万多千克的气压，平均每平方厘米（手指印那么大的面积）也就1千克左右。

适当的肌肉锻炼可以使致密纤维能横向分摊打击力，能起到蹦床那样的可以将打击力分摊到更广域面积的作用。武林高手通过练习，让腹部肌肉更加结实，获得可以保护骨骼和器官的具有横向分摊作用的致密纤维组织，局部肌肉可以承受800磅的击打力。

气功与经络理论

气功练习者即使不对特定肌肉进行这样的练习，也能承受近500磅的击打力，能承受每秒22米左右的击打速度。尤其是锁骨上方的喉部软组织每平方英寸（约6.45平方厘米）可以承受860磅到2900磅的力。那么到底什么是气？

有许多关于气功中的“气”是什么的争论。除了呼吸循环系统中有空气进入，身体其他部位不会有足够的令气功师具有特异功夫的空气。当然，经过适当呼吸训练，在吸气后通过呼吸系统的膨胀，将腹部鼓起为圆弧形，能起到分摊击打力的作用。

气功师感觉到的“气”或者使用的“气”，可能是一种内压。身体肌肉是一种软物质，介于固体和流体之间。水中有水压，因此作为软物质的

肌肉也可以产生内压。本来，我们体内的正常内压可以抵抗大气压，要知道大气压平均每平方米有一万千克。

经过特殊练习，能在特定时刻产生多余内压不难理解。气功练习时，许多动作中有四肢和躯体的弯曲和扭转。这种卷绕锻炼，可能使肌肉中的纤维、血管以及毛细管等获得特定有序的排列，一些与血流、淋巴输运和其他液体输送相关的细微管道可能更通畅，伸缩性更强，朝着更能激发内压和朝特定位置与方向传递内压的方式排列和变化。获得锻炼的内脏也可能通过畅通的血管或呼吸配合输送内压。

身体内液体的输送速度一般极慢，但压力波是一种振动，传播极快，比空气中每秒340米左右的声波传递速度还快数倍。例如，水中的压力波传播速度为每秒1.5千米左右。不知，气功中的意念发功，是否是在控制一些组织或器官的运动，激发可迅速到达身体某一部位的压力波。

除此之外，压力波到达端点时，不会无缘无故消失，会反射，就像乒乓球在台上反射一样。反射的压力波与入射的压力波的叠加有时会导致相互加强。中医把脉的脉搏点在手腕，即上肢的末端位置，跳动明显。心脏产生的压力波如果是在那里反射、再叠加，那么压力波的振幅应该可以在那里得到加强。心跳频率比1赫兹高一点点，每分钟约75次。假设心跳引起的压力波在血液传播，血管长度是1米。如果依然用水中的压力波速度来计算，那么从心脏到达血管末端只需要不到1毫秒，一个来回就是1毫秒左右。或者说，压力波在血管中的来回反射的频率是1000赫兹量级。这比心跳快多了，因此不会和心脏一起共振。

中医认为人体上有经络，经络上和经络外有穴位。穴位确实存在，而不像一些质疑者否认的那样。例如，盐城师范学院、第二军医大学、

复旦大学和上海应用物理研究所的刘成林（音译，Liu Chenglin）等九名科学家，于2014年在《电子光谱学和相关现象杂志》（*Journal of Electron Spectroscopy and Related Phenomena*）报道了经现代仪器（如CT扫描技术）探测的穴位特性。他们发现，在穴位所在的位置，不仅微量元素比别的区域多，而且微血管更细更密集（比非穴位区可以密集两到三倍）。

古代无法将穴位和身体内部的功能器官的链接关系用现代科学语言描述，因此统称为“气”。既然穴位上的微血管更密集，那么至少表明穴位有一些功能作用。按中医理论，不同的穴位（通过按压或者针灸）对不同症状有缓解甚至治疗作用。如果穴位是身体功能器官的特定排污口，那么当身体出现某些症状，对应的排污穴位可能排污不及时而出现堵塞，于是该穴位按压时可能出现胀痛或红肿。也可能出现某种类型的临时损伤，于是按压时出现刺疼感。

恰到好处的手指按压或者针灸可能有助于疏通被堵塞的穴位，甚至修复临时损伤的穴位，让排污畅通，从而达到缓解或治疗对应症状的目的。虽然如此，每个穴位如何与对应功能器官相连不是特别清楚（也许就是不太明显的血管）。穴位除了可能是特定的排污口，也许还存在极其细微难以探测到的神经组织、内压传输组织或韧带一类的牵引组织与某个功能器官相连。如果有这些功能，那么刺激时也就把刺激的作用传输到对应的功能器官之中。也许，有一些穴位是一种阀门或开关类的局部组织，被刺激时可以激发具有不同频率的神经信号或内压波（如同笛子按住不同风口能激发不同频率的声波一样）或其他牵引信号，而身体不同的功能器官可能对频率和振幅或其他牵引信号有种选择。

经络理论实践了数千年历史，不会违背科学原理，只是目前尚未完全按现代科学方式了解透彻。尚不了解并不表明不科学。

2. 生长与衰减 动物大小 技术能力的演化

数目或尺寸的增长或衰减是非常普遍的过程现象。技术的演化、癌细胞数目增长、头发与指甲的增长，都是生长问题。例如，手指甲可能每个月增长3.47毫米左右，脚趾甲每月增长1.62毫米左右，头发每月长长1.25厘米左右（这些数据可能因人而异）。有生长就有衰减，连构成物质的原子都可能衰减。技术的演化也像生命的演化一样，也有某种规律。尤其是一些重大突破，该出现时，会同时被不同的人想到。好像是技术发展的自身规律决定了技术的突破和重大创新的出现。倍增现象和半衰现象分别都是一枝独秀的增长和衰减。流行病患者数目、生物的数目等演化现象，同时被增长与衰减两类因素左右。这如同人的成长既会遇到正面影响也会遇到负面影响。两种相反因素的博弈，结果要么此消彼长，要么势均力敌。

倍增时间 癌症（图1.34）

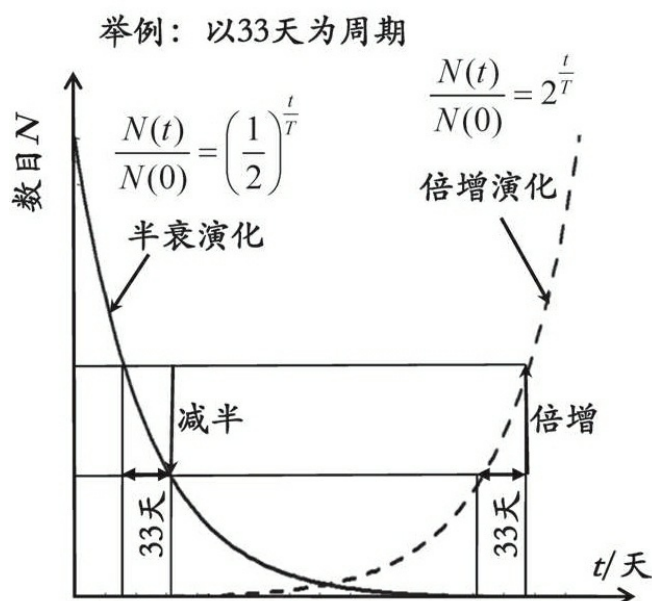


图1.34 倍增与半衰的数目随时间变化

技术能力或者生物的数目可能会快速生长，往往用倍增时间衡量生长快慢。数目翻倍需要的时间是倍增时间。技术上一个著名的例子是摩尔定律，即半导体芯片上集成的晶体管和电阻数量将每隔18个月左右（也说24个月）增加一倍。这18个月就是倍增周期。摩尔定律不是一个严格的结论，但一直用于指导芯片发展趋势。往往以此来说明传统计算机的运算与存储能力的发展有多迅速。

倍增时间可以用于预测疾病爆发的时间。例如，癌症细胞由正常细胞衍生而来。增值过程中，癌细胞1个变2个，2个变4个，……。胃癌、肠癌、肝癌、胰腺癌和食道癌的倍增时间平均是33天左右；乳腺癌倍增时间是40多天。数目达到10亿个以上我们才会察觉，从第一个癌细胞开始以倍增速度增长，3年左右就会达到10亿左右。达到10亿后，只需要30多天就达到20亿。因此，癌症一经发现就可能到了晚期。由于倍增原理，癌症越到晚期进展得越快。

癌症由于发病率越来越高，因此人们越来越恐慌。其实，人均寿命在增加，年龄越大可能更容易得癌症或者说寿命的增加越容易熬到癌细胞扩散从而被发现。过去由于人均寿命短，一些现在能轻易治疗好的疾病，在过去可能就夺走了生命，挨不到癌症的出现。于是给人印象癌症患者越来越多了。

癌细胞虽然是病变，但免疫系统却把它们当正常细胞，不予攻击。肿瘤先是在原位生长，接着游离原来的器官，循着淋巴液转移至附近的淋巴腺，最后是远距离转移。有关癌症治疗总是有激动人心的报道，人类翘首以盼能彻底攻克癌症。人们喜欢有某种方式不让它出现，不让它长大，不让它扩散，不让它造成破坏。一旦发现癌症，要么杀死它们，

要么清除它们，要么饿死它们，要么通过免疫方式抵抗它们。

由于目前的医疗水平尚未达到完全有效的程度，因此总是善意地提出一些对癌症预防可能有作用的非医疗类建议。如拥有开朗的心态、乐善好施的行为、规律的作息时间、戒烟戒酒、健康的饮食。这些措施也许有助于避免癌细胞的原位生长。由于局部转移需要循着淋巴液，因此中医养生中有通过刺激淋巴反射区来缓和癌症转移的建议。癌症患者由于癌细胞吞噬，因此身体虚弱，因此在一些足疗著作中，也有刺激脾反射区应对癌症的建议。如果出现远距离转移，在目前看来，也许手术、化疗等现代手段是延长存活期的最有效方式。

衰变速度、半衰期与年代测定（断代）方法 (图1.35)

有的演化是数目不断翻倍，有的则是数目不断减半。与倍增时间对应，减半所需要的时间是半衰期。由此依据某物体在死亡前吸入的不同元素的比例的当前值，可以测定该物体的死亡年代，包括化石年代。

一些原子具有放射性，其原子核不停辐射一些粒子或射线（主要有 α 粒子、 β 粒子、 γ 射线或中微子），从而转变成其他原子核（如碳14转变成氮14）。这一过程叫衰变或蜕变。放射性元素的原子核有一半数目发生衰变时所需要的时间，叫半衰期。碳12没有放射性，不会衰变，但其同位素碳14则会衰变，半衰期是5730年。

大气中有一定数目的碳12和碳14。大气中两者数目的比例不会变化，因为有宇宙射线来维持这个比例。碳12和碳14与氧结合成二氧化碳，进入所有活组织，先被植物吸收，后被动物吸收。故动植物死亡

前，碳12与碳14的比例与（过去和当前）大气中的两元素的比例是一样的，且大气中两种元素的比例基本不随年代变化。动植物死亡后，不再吸收新的碳，而其中碳12不衰变，碳14衰变。故化石中的碳14与碳12的比例随年代按半衰期5730年来变化。用现代工具测量二者的当前比例后，就可以算出动物死亡的年代。

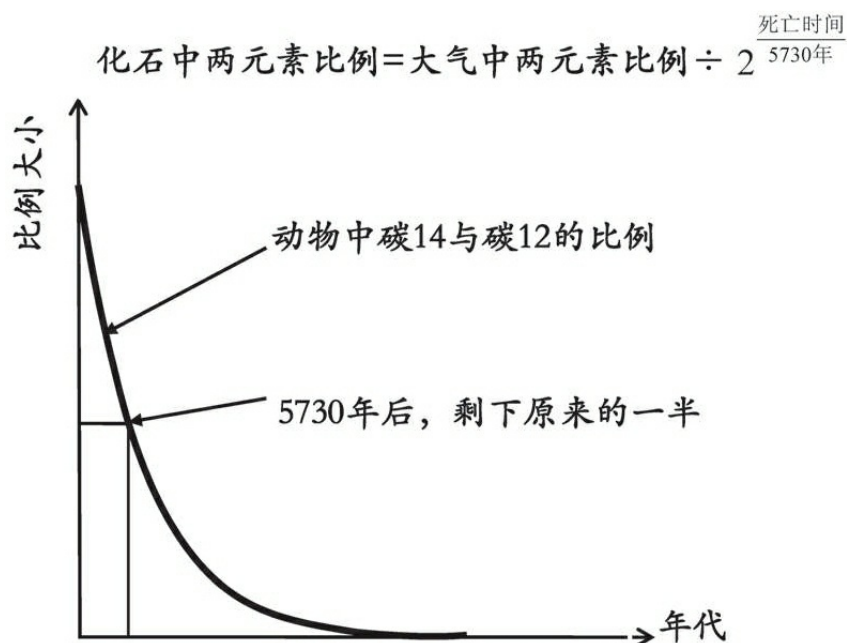


图1.35 化石中碳14与碳12比例变化

由于碳14的半衰期只有几千年，因此死亡太久的动植物，剩下的碳14比例太少，很难测准，因此存在一定的误差需要校验。铀238半衰期为45亿年。含铀的化石，其铀在衰变中产生铅206。通过测量含铀的化石中剩下的238铀和206铅的含量的比，就可以较为准确地计算出该化石的绝对年龄。

科普定律：动物的体积越来越大 我们会成长成巨人吗

正是因为有了测定年代的方法，科学家能从化石测出动物死亡的年代。依据化石反映的动物的大小，就可以得出动物的体积是越来越大还是越来越小的结论。

19世纪，美国古生物学先驱，爱德华·科普提出了一个定律，后来称为科普定律。该定律认为，动物起初往往体型较小，通过不断进化，体型会变得越来越来大。科普定律一度被认为不是一个自然恒定规律。为了验证其正确性，斯坦福大学的佩恩和海姆利用化石记录研究了动物体型进化的趋势。他们的结果表明，科普定律是正确的，于是进化可以被预测。

在过去5.42亿年里，海洋动物的平均体型增大至原来的150倍。例如，美国西海岸温暖的浅海中的海胆的正常体长，200万年前是5厘米，现在增长到了30厘米。这些结论涉及了17000个海洋动物化石，占到了有化石记录的海洋动物总数的75%，涉及的种类几乎占地球史上所有动物种类的2/3。

科普定律与我们有关身高的常识也很吻合：与父辈的身高相比，新一代身高在增加。然而，我们会进化成庞然大物吗？

动物体积以及数目的演化也受能量守恒定律和代谢原理的限制。克莱伯法则指出，动物的基础代谢率是体重的 0.75 次方。基础代谢率是在清醒而又极端安静的状态下，不受肌肉活动、环境温度、食物及精神紧张等影响时的能量代谢率。基础代谢率越高，消耗的能量也就是食物会越多。因此，体重越大，基础代谢率就越大，消耗的食物就越多。

由于一个生物圈上的能源和食物是有限的，因此，单位地域面积的动物数目（即数目密度）随其对能源或食物的需求增加而下降。也就是

说，重量越小的动物，数目密度就可以越高。蚂蚁和白蚁远多于人类，因为它们的体积小多了。人的数目不可能比蚂蚁和白蚁多，否则太阳提供的能量不足以养活人类。或者说，数目一定的情况下，体重不可能无限制增长。

弗格斯·辛普森认为可居住的星球的半径应该比地球半径的1.2倍要小，宜居星球的大小更接近火星大小。弗格斯·辛普森依据上面那样的分析，认为地球外的高等生物的体重应该在300千克以上，比现代社会迄今纪录到最高的人，2.72米的美国人罗伯特·瓦德罗（Robert Wadlow），还要高。因此，未来地球人是否体重也高达300千克以上，不得而知。

自然界中，动物数目的演化受制于对食物链的竞争以及环境变化对食物链的影响。对于人类，社会发展和宗教信仰也对不同的人群数目的演化产生影响。如果维持一对夫妇生两胎，那么总的人口数目就不会有太大变化（寿命的增长、灾害和战争也会导致总人口数目的变化）。研究表明，为了使一种文化得以保留25年以上，生育率必须维持在一对夫妇生2.11胎以上。如果生育率低于1.9，那么文化消失现象很难逆转，进一步如果低于1.3，那么根本不可能逆转。如果一对夫妇平均生4胎，那么人口就会出现以代为周期的倍增现象。如果只有1胎，那么就会出现以代为周期的半衰现象。随着社会的进步，生育率呈现越来越低的发展趋势。据报道，2007年，法国生育率为1.8，英格兰为1.6，希腊和德国为1.3，意大利为1.2，西班牙为1.1，欧洲共同体31个国家的平均值为1.38。近期欧洲人口靠移民来维持数目。受此影响，人的体重也许增长得更快。

技术像生命体一样演化

科学技术的发展，往往被认为只是少数天才的科学家凭借与生俱来的天赋、卓越非凡的洞察力和坚韧不拔、不屈不挠的毅力推动的。然而，美国科技哲学家，《连线》（*wired*）杂志共同创始人凯文·凯利在《科技想要什么》和《失控》等著作中则认为科技的演化具有生命的特征，即科技有自身的演化规律。一些重大发现和技术突破似乎被写在了历史的轨迹上，只是由谁来完成而已。

凯利认为，地球上的生命包括七种基本形态：植物、动物、原生生物、真菌、原细菌、真细菌、技术。技术的演化和前六种生命体的演化相似，因此技术也可以理解为是一种生命形态。

如同生命的演化，技术的发展会以先前的技术作为基础。以植物为例，合适的土壤、柔和的光照与风调雨顺是植物生长的条件。一片土壤有了这种生长条件，不是出现树木就是出现花草。一树种是否得以蔓延带有偶然性，但大量的树种出现优胜劣汰的发展则是必然。

同理，如果技术有了萌发的土壤，求知欲与经济、市场、竞争等驱动因素会使大量科技人员去钻研与研发，不是这里获得突破就是那里获得突破。比如说，有了望远镜技术，求知欲自然会驱动人们去探索太空，这颗星球、那个星系自然有人发现。自然有人会观察到星系旋转。自然有人发现旋转速度异常，从而有人想象用暗物质去解释异常现象。现代通信技术、计算机技术、交通技术、医疗技术已经不满足人类的更高要求，于是驱动人们探索量子技术、无人驾驶技术、微纳米机器人技术。于是，这方面的技术自然会得到突飞猛进的发展。

牛顿时代，建立微积分既有了基础也有了驱动，因此牛顿与莱布尼茨可能同时想到。那时，微积分该出现了，即使没有牛顿和莱布尼茨，其他数学家也会发明。爱因斯坦时代，许多理论和实验结果均指向相对

论出现的必然性，因此无非是爱因斯坦先建立而已。爱因斯坦说过，如果他不提出狭义相对论，那么也会由别人提出。揭示重力之源的希格斯机制，于1964年被三组不同科学家分别独立提出，其中包括于2013年因此获诺贝尔物理学奖的恩格勒和希格斯。

科学发现与技术进步会通过文献出版、会议交流以及通信等方式得以扩散，这使发达程度接近的不同国家的科技水平在一定的历史时间段差异不大。电子出版尤其是资料免费的开放获取，使现代科技知识的扩散非常迅速。谷歌技术总监库兹韦尔甚至认为，技术的力量正在以指数级的速度增强，人类正处于加速变化的浪尖上。

在经济、军事、政治和自由探索精神的驱动下，各地科学家和发明家应会被驱动或者自觉探索基于先前技术的问题。这一方面导致创造被独立地由不同人提出，另一方面导致技术的发展更像一种自然进化，甚至技术的发展可以被预测，至少被部分预测。例如，摩尔定律就非常准确预测计算机芯片存储能力的发展。

既然技术演化具有生命力且遵循生物学演化规律，那么微积分、相对论、希格斯机制、计算机与互联网等就是时代必然出现的产物。

增长与衰减的博弈（图1.36）

倍增现象和半衰现象分别都是一枝独秀的增长和衰减。流行病患者数目的变化、生物的数目演化和城市与人口的扩张等，同时被增长与衰减两类因素左右，如同人的成长既会遇到正面影响也会遇到负面影响。两种相反因素的博弈，或此消彼长，或势均力敌，演化过程随着两种影响因素的消失而消失。

流行病的扩散与消失，小孩头发数目生长而老人的减少，人体身高先增后降等，都是生长因素与衰减因素共同起作用的结果。例如，一岁前的小孩，身高每年增长10厘米以上，之后一直到10岁多一点点，每年增高量逐渐下降，降到每年只有5厘米左右。之后增长速度又开始变大，女孩到12岁左右每年增长8厘米左右，男孩14岁左右每年增长9厘米左右。之后，每年的增长量又开始下降，到了20多岁，我们的身高就不怎么增长了，老了后甚至开始下降。

流行病开始爆发时，得病人数会与可感染区域的总人数成正比。这是因为，人越多，接触病人的人就越多。病人越多，接触病人的人越多。这导致感染人数成倍增加，刚开始会爆炸性地增长。随着尚未感染的人数下降、已经被感染的人的恢复、离世以及人为的干预，住院人数的增长将出现一个拐点，即增速开始下降的时间点。最后住院人数会出现一个极大值，接着开始下降。

以2003年3月份爆发的北京“非典”疫情为例，由于社会干预、出院和可感染人数下降等衰减因素，4月27日左右出现拐点，5月15日左右出现极大值，最后疫情慢慢消失。

疫情先增后减所满足的演化规律具有一定的普适性，包括人口增长率、城市扩张速率和环境污染的变化等可能满足相似的规律（数学上可用对数正态分布规律描述）：数目以及数目的增长率先爆发式增长，接着增长率出现极大值即出现拐点，随后数目达到极大值，最后数目开始减少。

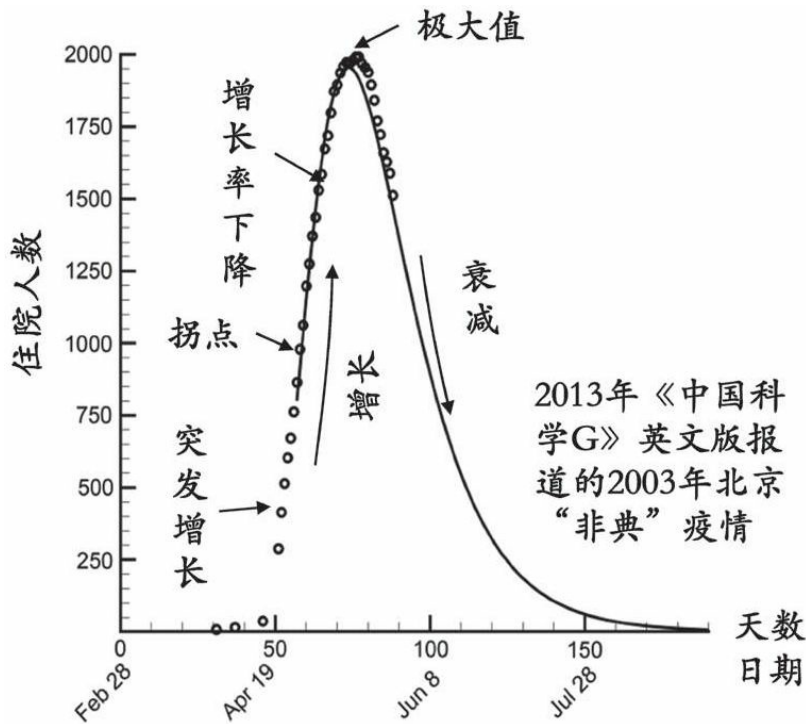


图1.36 先增长后衰减现象

生长因素促使增长，耗散因素促使衰减。与前面介绍的倍增或半衰只有单一因素不同，这种此消彼长或势均力敌的增长与衰减，是两种因素博弈的结果。首先是非常快速地增长，增长速率越来越大。流行病之类的增长太快的话，就会出现多种因素的干预，阻止疫情无限制地恶化。这种干预就是衰减因素之一。于是增长率会在某一刻达到极大值，而不会无限制地增长下去。增长速率达到极值的点，就是拐点。即拐过去后增长速率就下降了。拐点过后，还在增长，只是增长速率越来越小，一直趋于零。不再增长后，数目就达到了最大值，即增长因素与衰减因素势均力敌的位置。随后，衰减取代增长，数目从整体上开始衰减。

除了这些好理解的例子，自然界还存在许多类似的先增长后衰减的例子。具体到每个现象，都有具体的原因，但作为一个普遍现象，又似

乎存在更深刻的甚至统一的原因。

3. 演化的普适规律 省时省力省料

演化经常是如此的复杂，以致对每一类问题，细节性的过程需要大量专家去探索。虽然如此，演化的大方向和大结局至少可以用我们的常识和直觉去定性地理解和描述。演化是多种因素相互影响的结果，它们之间的相互制约会使表面上的对抗，隐含了某种相互包容以及妥协。这会使演化的方向、过程与结局遵循一些普适的规律。演化的方向也能体现削富济贫、缩小差异。演化的长远结局往往使差异最小化或利益最大化或者状态更稳定。这可能决定了演化对过程的选择：朝着缩小差异的方向，并构筑一些中间过程（如开水沸腾）或副产品（如生物圈的诞生、城市的出现），以便尽可能快且代价最小化地满足差异最小化、利益最大化和状态更稳定。省时、省力和省料不单是人类的智慧行为，大自然的演化也会这么奇巧。比如说蜂巢那样的六角形结构就是多方博弈的妥协，这种妥协就导致了省料，以及筑巢过程最快（即省时）。也因为如此，演化的终极结果往往可以预测，对演化规律的理解可有助于预测未来。

演化的趋向：缩小差异（图1.37）

高处的自由物体会下落，水往低处流，温度高的暖气片向房间内温度更低的地方输送热量。这都是朝着缩小差异的方向演化的结果。暖气片输热是在缩小暖气片温度与周围空气温度差异的过程。高处的水与物体，比低处的位置更高，势力（势能）更大。往低处流动与落下，也是与低处的水与物体缩小差距的体现。因此，演化的方向也可以说成是慷

慨解囊、削富济贫。缩小贫富差距，才能更稳定，这才是演化的方向。

如果把初始的差异定义成有序状态，缩小了的差异定义成无序状态，那么可以理解为演化朝着更无序的方向演化。把树叶茂盛当作一种有序状态，那么秋风扫落叶后，树叶随机散落在地面当然可以看成一种无序状态。落叶无序是自然的过程，树叶散落后不会自动跑回树上。在一杯清水中滴一滴墨水，墨水就散开了，即显得更无序了。散开后，在没有特殊干预情况下，浑水不会自动变回原来的一滴墨水 and 一杯清水。

在水中滴一滴墨水，变得更无序了



从污泥中长出的荷叶



图1.37 无序现象的出现

因此，让一个包含了差异的状态自然演化，会朝着更无序的方向演化。

表述无序的正规术语：熵

物理学上用“熵”表示无序的程度。熵越大表示无序程度越高，熵越小表示无序程度越低。因此有一个物理学原理是这样说的：让一个没有外部作用干预的孤立系统自然演化，其熵只会增加不会减小，即只能越来越无序而不能越来越有序。也就是说，孤立系统中的差异会缩小，顶多不变。

这是一个普适法则，物理学上把这个叫热力学第二定律（热力学第一定律是指，给一个系统注入能量，那么该系统吞掉一部分能量变成内部能量即内能，剩下的一部分给外部做点贡献即用于膨胀做功）。爱因斯坦认为热力学第二定律是具有普适价值的唯一物理学理论，即放在哪里都能用，绝对不会错。用热力学第二定律可以证明，不可能制造出永动机（据报道，量子行为有违背热力学第二定律的可能性）。

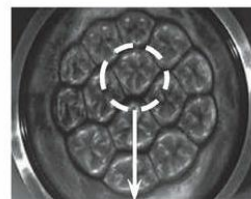
高温空气向周围低温空气传送热量，而不是反过来由低温向高温输送热量使暖气片越来越热。为何这种现象非得说成是变得越来越无序了即熵增加了？原来，空气之所以有冷暖，是因为空气分子在做杂乱无章的热运动。每个分子一会儿跑到这里，一会儿跑到那里。温度越高，那么这种热运动就越强烈，从某种角度上说就越无序。因此，暖气片的热流传到冷空气中，使冷空气变热了分子热运动加强了，就是变得更无序了，即温度的演化朝熵增方向演化。

当然，我们会说，地球上生物的诞生难道不是变得越来越有序了？尤其像人一样的动物，有形的外表、功能齐全且能完美协同的器官、各种循环系统，这难道不是一种有序的表现？布满空隙的藕条、雨伞一样的荷叶、灿烂的荷花、鲜嫩可口的莲子，这可是完美的有序结构，它们明明是从无序的污泥中生长而成。小荷才露尖尖角，漫无边际的莲花雕刻着清澈的湖面，这显然是一种高度有序的自然画卷。其实这不违背热力学第二定律，因为该定律是指孤立系统。地球生物圈接受太阳光等外力的干预，不是孤立演化出来的。再说，眼光放大点看，相比于只有炙热的太阳和光秃秃的行星的太阳系，生物圈的存在就是一种无序。

省时省力 最大熵增率 最小作用原理（图
1.38）

如果说差异朝着缩小的方向演化，那这个普适法则却并没有告诉我们以什么样的速度演化。大自然也像人类的某些行为一样，有点着急，有点偷懒，以致由一个状态向另外一个状态的演化，会寻求最省时和最省力的演化过程。费力不讨好的事情，拖拖拉拉的事情，连大自然都不干。当然，科学家不会以这种方式来表述，把熵拉进来说事，就会显得科学一点。

水烧开后，锅底产生热对流元胞，称为贝纳尔涡



热水从元胞中心线上升，冷水从侧边下降

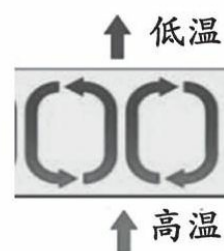


图1.38 沸腾的水用元胞加速换热

加热平底锅一薄层水。锅底的水在加热后，温度增高，甚至变得更轻。热量从锅底高温位置，通过薄层的水散到水的上表面。这种在静态水（空气中类似）中，热量由高温位置向低温位置传播的过程就是热传导。继续加热，就可能出现许多六角形元胞（也称为贝纳尔涡，又称为瑞利-贝纳尔元胞，因为是他们通过实验研究提出了形成机理及其影响因素的）。六角形中心的水在上升，上升后的水沿六角形边缘下降。这种由于加热后引起的在浮力作用下的水流（或气流）上升，也称为自然对流。在大气中，在高度越低温度越高的区域中，也存在这种“沸腾”导致的对流现象。你在空中释放一根小毛毛，小毛毛可能往上蹿，这就是自然对流在把小毛毛举向高空。

以上现象，以及非常多的其他演化现象，蕴含着两大基本原理，按科学语言就是最小作用原理和最大熵增率原理，实际上就是省力和省时原理。这些简单普适的原理，左右着所有运动与演化。最小作用原理的提出者皮埃尔·路易·莫泊特斯（PierreLouis Maupertuis）认为，自然界尽

量节省其作用付出的代价。有速度的物体在没有外力时，走直线，因为走直线的代价最小。有转动因素驱动而生成的台风和澡盆涡，近似以圆周路径旋转，因为只有这样，各点走的路径才基本相似，否则就会相互干扰。能一样就一样，避免因差异带来的冲突，也是省力的一种表现。

齐格勒（Ziegler）提出的最大熵增率原理是指演化速度需要遵循的规律。虽然表述的是无序化的速率要最快，但也可以说成是省时原则。加热平底锅薄层水的过程中，出现六角形元胞，就是让热量的输送最快，或者说缩小上下温差的速度最快。如果说缩小温差就是一种熵增，最大熵增率原理就是指演化使熵增加的速率最快。

尽快完成，代价最小化，这就是自然所遵循的普适法则。如果局部出现有序结构（如美丽的生物圈，流线一样的河道，圆圈型的涟漪，等等）有助于满足整体上代价最小化和最大熵增率原理，那么就会出现局部有序结构。弯弯的河道尽量光顺，才能使水流顺畅，才能使高水位更快地流向低水位，使水位差缩小的速度最快，反映位势差的缩小速率的熵增率最大。

生物圈的形成与优美的演化，表面上看，局部或者个别生物本身更有序、更优美，从而熵更小。但把地球看成一个整体，其中的生物部分变得有序、自身的熵最小，则整个地球产生的熵最大化。

神秘的六角形（省料）（图1.39）

煮水形成的瑞利-贝纳尔元胞中，温度增高的水通过浮力在六角形中心上升，上去后将热量传递给空气从而冷却了一些，又在旋涡带动以及重力作用下，迫不及待地沿六角形边缘下降，尽快去吸收新的热量再

从中央上升。这好像是人为设计的尽快向上搬走热量的机器，没有什么其他手段比这效率更高。

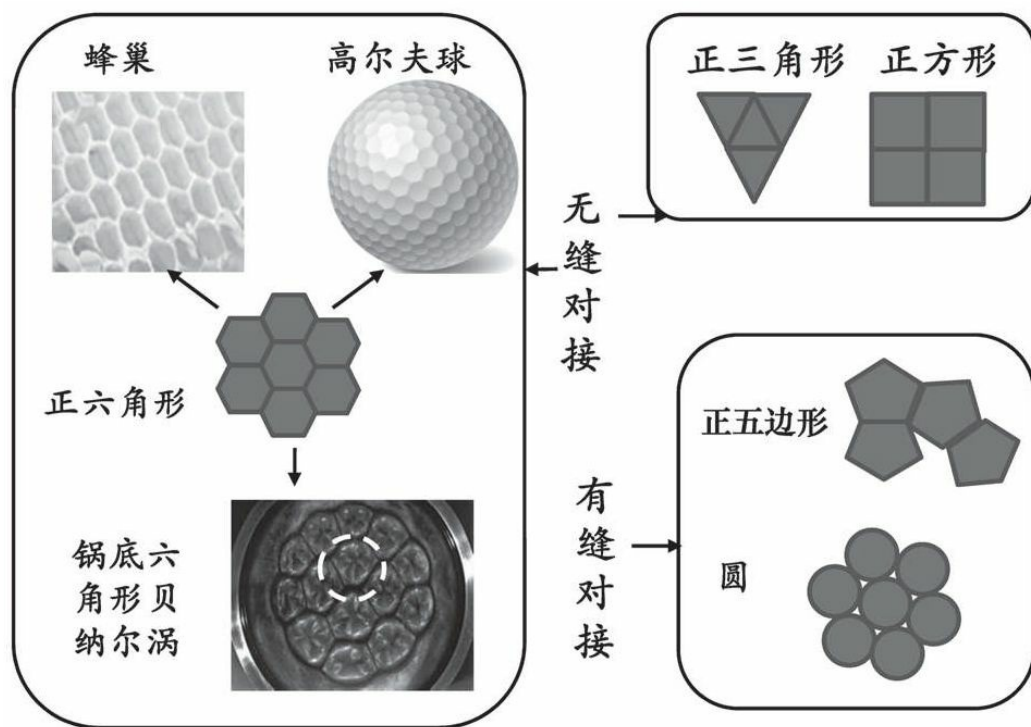


图1.39 能无缝对接的六角形最省料、围的有用空间最大

我们会问，为何是六角形元胞，而不是圆形。原来，六角形可以挨着一个接一个，而圆相接时，三个相邻的圆之间会留下空隙，不能最大化地覆盖锅底。正三角形、正四角形和正六角形都能无缝地铺满整个平面。其他规则多边形都不能，圆也不能，因此不能有效利用全部空间。六角形与三角形和四边形相比，同样面积情况下，边长的和最小。因此六角形既能占满空间又最省料。

由数千个巢房组成的蜂巢，每一个都采用正六边形，可以无缝覆盖，最大限度地利用空间，利用了六角形所需材料最少且可使用的空间最大的原理。当然，蜜蜂的聪明可能并不是因为它们懂数学，可能是在巢穴中相互挤来挤去，最后挤成六角形，就没法再挤出别的形状了。小

蜜蜂各自占据自己的巢穴，都想自己的空间最大化，于是应该会挤来挤去。最后的结果是妥协，使各自的空间最大，总的空间不变。利益最大化，却最省料。

高尔夫球的酒窝凹槽是为了打乱气流，减弱拐弯涡带来的压阻，以便飞得更远。为了最有效地打乱气流，也可能做成六角形的。

冰的分子具有六角形结构，而生成冰的水则是三角形结构。显然，这会使冰比水轻。这种六角形结构还使角点更容易抓捕小水珠，长出六角形雪花。

4. 演化规律能解释我们的世界吗？

用某种眼光看，运动无非是平动、转动和振动，以及这些基本运动的叠加。但把许多物体运动构成的随时间的演化一起看，就包含了一些普适原理，如省时省力省料原理。这很容易帮助理解开水沸腾等现象，甚至能帮助理解旋转的台风和河道的演化。但我们的世界包含的东西太多，我们能用演化的普适规律去解释吗？尤其是演化的结局是什么？是朝着简单、简约和简美的方向演化吗？

我们的世界灿烂无比（图1.40、图1.41、图1.42）

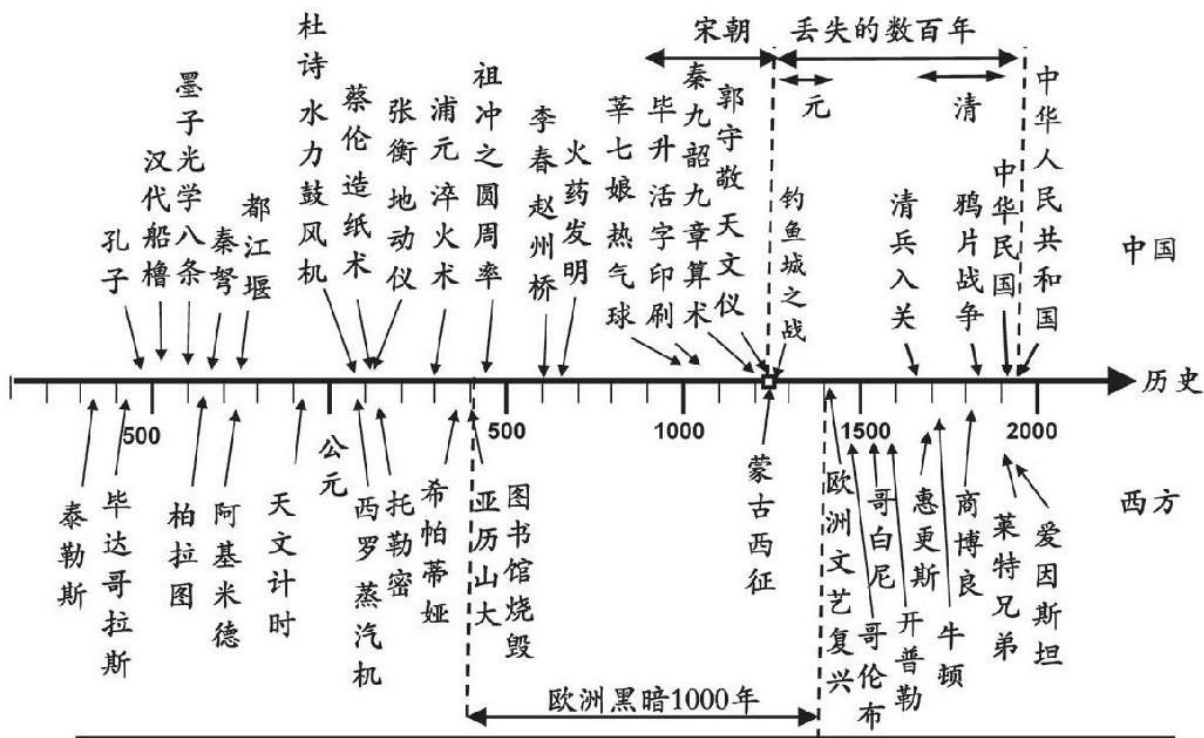


图1.40 在时间轴上中西科技文明的历史演化



图1.41 英国科学家绘制的部分国家女性平均脸型



图1.42 预测方法包括机会型、经验型、科学型和综合型

演化的普适原理不知能否用于解释千变万化的整个世界。地球包含万物，我们用世界一词来囊括一切。与其用数字来描述地球或者世界，我们还不如用文字。也许，文字中就包含了不适合数学的数字。大到无法准确度量，小到无法精确测量，远到除了无限否则无法描述，多到除了无穷否则无法计数，神秘到除了遐想否则难以表述。一些无法解释的奇特现象要么被当成精神寄托，要么被视为迷信，要么用未知科学进行关联。甚至诸如灵魂之类的敏感问题，也试图用诸如量子纠缠等现代物理学知识去解释。

地球上有大气、陆地、海洋，陆地上有江河、湖泊、山丘、峡谷、瀑布。多样的生物，包括植物和动物。当然，更有我们人类。人类的意识、智力、长相、思维、行为……，以及人类创造的建筑、科技、文化、艺术和人际与国际关系，等等。从天空看地球夜色，那是建筑物与

光的岛屿，因为黑暗而灿烂。

有时，一些历史重大事件可能导致不同区域与不同国家的相对发展速度出现变化。公元四世纪，代表当时科技文明的亚历山大图书城的烧毁，使欧洲进入黑暗一千年，中国继续向前发展，使中国科技一度领先西方。随后的欧洲文艺复兴导致近代科技文明向前发展，而与此同时，中国进入了有百年停滞不前的时代，几乎正好与欧洲黑暗一千年前后错开，这导致本来领先的科技文明出现落后。

有的已经成为历史、有的继承到现在。有的正在创造、有的属于未来。还有的属于想象、梦境、期盼、意念。或永远留存，或瞬间即逝。或流芳百世，或遗祸千年。或在诞生之初，或在毁灭之末。有的是实在的，有的是虚拟的。有的，属于概念与设计，或属于痴心幻想，或导致重大突破。


文字、画、传说和古墓，记录着我们遥远过去的点点滴滴。近代的记忆，补充了照片与视频。望远镜让我们看到了远方。我们开始用色彩表示科学数据，照片与图片成了科学与艺术的桥梁。于是，我们有了虚拟世界，科幻世界。遥远的过去，要么成了古董、要么成了典故、要么成了史书、要么成了传说。继承过去是为了充实现实，珍惜现实是为了面向未来。过去与今天的一切都是通往未来的基础，向往美好的未来是演化的动力，这种向往驱动我们实现更美好的明天。继承过去并不等于厮守过去，而是为了获得更好的未来，就像科幻电影展示的那样。

共性、多样性、差异性，和谐与对立，使世界如此灿烂、如此美好、无穷无尽。单单就我们人类，居然分成了埃塞俄比亚人、奥地利人、墨西哥人、巴西人、澳大利亚人……当然还有我们中国人。这是地理、文化、政治、经济和军事等综合演化的一种格局。各国之间人的平

均的长相不一样，每个国家不同人长相不一样。肤色不一样，文化不一样，习俗不一样，制度不一样。文明在趋同，但文化却存异。

也许，这些也可以用大的演化规律来理解！对演化规律的理解并不总是为了解释过去和描述现在，有时是为了预测未来。预测有机会性、有经验的作用，但也有科学的指导。综合地运用这几类预测技术，可以提高预测的准确程度。

汉字的演化（图1.43）



大世界的演化那么复杂，哪怕是文字的演化也是如此。演化的结局是复杂无比，还是变得简单、简约和简美？

我们这里不去讨论汉字形态的历史演化，而只看笔画。造汉字，笔画如果太少，就组成不了几个字，因此笔画少的字少的可怜。一笔的字也就是“一”这个字。两笔的字大概有“二”“十”“八”“人”之类的，十个左右。

于是，笔画越多，越能造出更多的字。可是，一个字的笔画如果太多，就记不住了或者书写起来太麻烦了。笔画少了造出的字不够多，满足不了要求，因此希望笔画越多越好，这是促使笔画数目增长的因素。笔画多了难写，记不住，这是促使笔画数目增速减少的衰减因素。最后造多少字，以及这些字的笔画多少，是这增长因素和衰减因素博弈的结果。

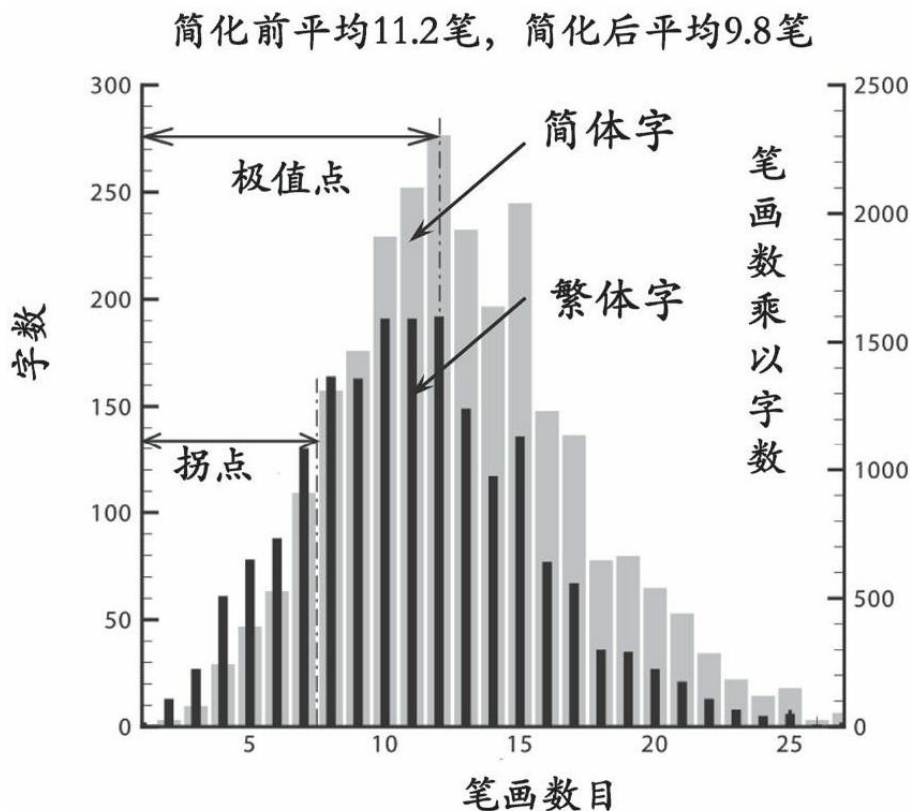


图1.43 不同笔画的常用汉字数数目

期刊《文字改革》在1958年第3期发表了“常用汉字笔画统计”数据。结果表明，汉字简化前，12笔的常用汉字数目最多。简化后，10笔的常用汉字最多。笔数超过12笔或10笔，常用汉字的数目就开始减少，到了26笔，就差不多没有了。

有一个特殊情况，笔画数目从7向8过渡时，汉字数目变化最多（简体字由180个变到205个；繁体字由130个变到164个），再往上，相邻笔画数目的汉字数目变化就下降了。似乎7是个障碍。这个障碍难道是巧合？

5. 数字7

汉字7笔以上就有障碍了。常见的能量形式为机械能、热能、电能、磁场能、光能、化学能和（原子）核能等7种。基本物理学单位正好有7个。数字用于表示计量、顺序、大小、位置、程度、快慢和变化等7个方面。还有我们熟悉的7个音阶、七言诗、七色光。难道数字7隐含了什么秘密？当然，一周有7天、有7颗北斗星、地球有7大洲，人们认可7大奇迹也许只是一种巧合。对7的某些偏爱引起了人们的思考：是否存在深刻的自然或数学或心理学原因，还是纯粹属于巧合？这个问题还吸引了不少科学家的关注。对数字7的钟爱，也是演化的一种结局。

数字7的普遍性

在宗教仪式、医术、节假日、图形与符号、教条、地理分布、管理、关系等方面涉及数字7的实例非常多。与时间相关的7：一周7天、人生7个阶段、圣经中上帝7天创造世界、七夕情人节。奇迹、难题、生命与自然：世界7大奇迹、克雷数学研究所发布的7大数学难题、7种生命形态（植物、动物、原生生物、真菌、原细菌、真细菌以及技术）、海洋波浪的波长超过浪高的7倍就会破碎。地理：地球7大洲、19世纪中国分为7个省、战国七雄。星球：北斗七星、日月金木水火土7颗星球。宇宙：卫星、行星、恒星系、星系、星系团、超星系团、宇宙。艺术与色彩：七言诗、7个音阶、七色光（红、橙、黄、绿、青、蓝、紫）。寓言与故事：竹林七贤、建安七子、杨家七将、七仙女下凡、七夕鹊桥会、七个小矮人、辛巴达七海传奇。我们的日常生活离不开柴、米、油、盐、酱、醋、茶。

我们还可以举出更多的例子。

（1）物理学7个基本单位（图1.44）

物理学包含许多规律，却只有7个基本单位：长度（米，m）；质量（千克，kg）；时间单位（秒，s）；电流强度（安培，A）；热力学温度（开尔文，K）；物质的量（摩尔，mol）；发光强度（坎德拉，cd）。其他单位可由它们导出，如速度为长度单位除以时间单位。基本单位以外有两个辅助单位：弧度（rad）和球面立体角（sr）。



图1.44 科普作者多米尼克·瓦里曼（Dominic Walliman）的物理学地图

（2）7种常见味觉（图1.45）

7种味觉是酸、甜、苦、辣、咸、鲜、涩。其中，酸、甜、苦、辣、咸这5种味觉似乎全球通用，其他几种味觉不同国家稍有差异。比如说，欧美将味觉大致分为酸、甜、苦、辣、咸、金属味、钙味。

不知是鲜味和金属味两种味觉的差异导致了餐饮中分别进化出了筷子文化和刀叉文化，还是反过来，因为我们使用筷子从而进化出了鲜味味觉，西方人因为使用刀叉而进化出了金属味觉。

五味调和百味香，即不仅仅是这几种孤立的味觉，它们合适的调配能产生馨香可口的美味，如同不同音调的声音调和在一起可以形成美妙动听的音乐。调音，可以让人听得悦耳，调味，可以让人饥饿。

（3）7组脸部表情（图1.46）与7种情绪

愤怒、快乐、恐惧、惊讶、厌恶、悲伤、轻蔑。政法大学微表情研究小组公布了更详细的说明。其他数千种表情可以看成7组基本表情的组合，如同连续变化的颜色可以看成7种基本颜色的组合，以及音乐可以用7个基本音阶组合一样。

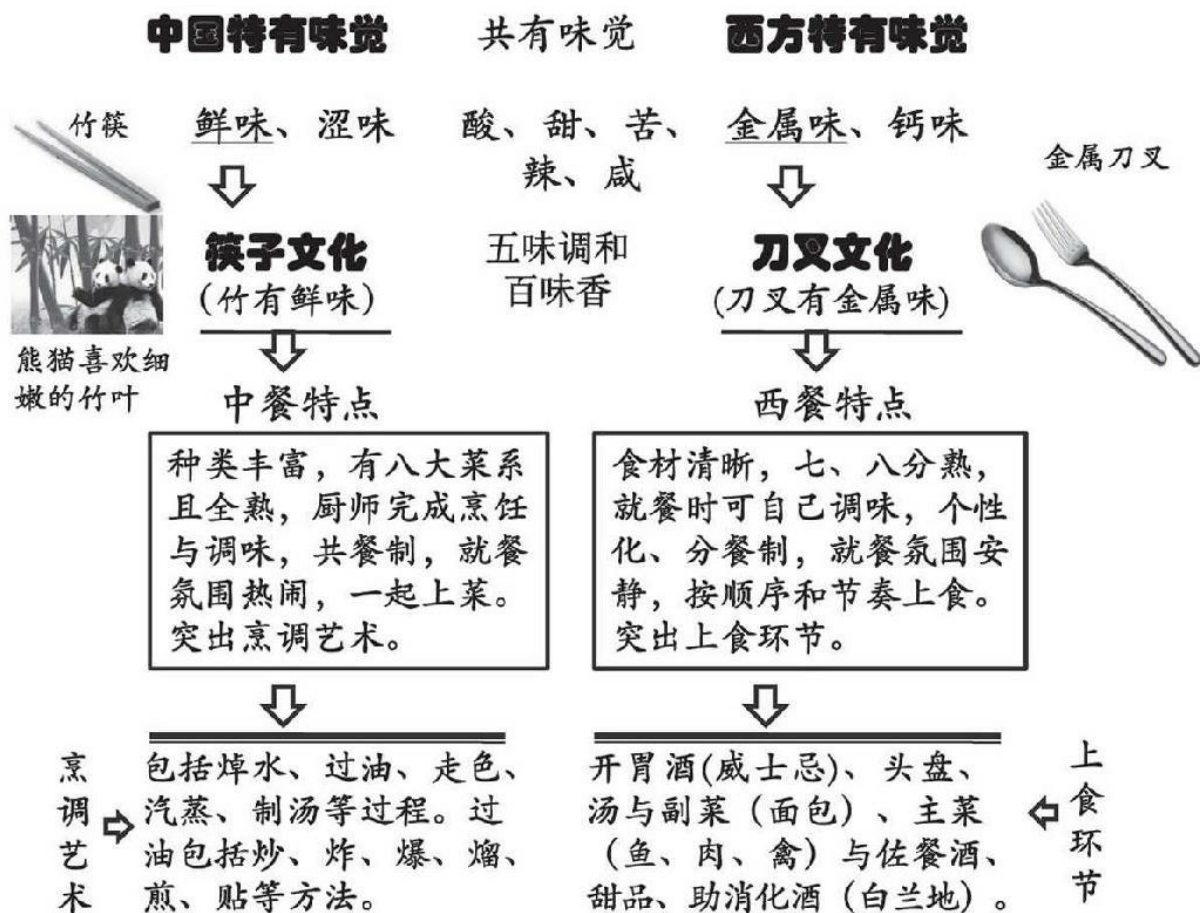


图1.45 七种味觉与中西饮食文化



图1.46 面部七种基本表情

七情六欲中的七情是喜、怒、忧、思、悲、恐、惊。温和的情绪一

般不会使人致病，突然强烈或长期持久的情绪刺激就会超过人体本身正常生理活动范围，使人体机能紊乱，脏腑失调，导致疾病的发生。五脏六腑中的五脏是指心、肝、脾、肺、肾。六腑是指胆、胃、小肠、大肠、膀胱和三焦。按中医理论，怒伤肝、喜伤心、思伤脾、忧和悲伤肺、惊和恐伤肾。

（4）7组经络（图1.47）

中医认为身体上有经络和穴位（足部有反射区），经络理论被广泛用于养生甚至（辅助）疾病治疗。

人体有14条经络，被归入7个组，分别是任脉 / 督脉、肾经 / 膀胱经、肝经 / 胆经、脾经 / 胃经、心经 / 小肠经、心包经 / 三焦经、肺经 / 大肠经。每条经络上穴位数目不一样（初步统计，任脉24个 / 督脉28个；肾经27个 / 膀胱经67个；肝经14个 / 胆经44个；脾经21个 / 胃经45个；心经9个 / 小肠经19个；心包经9个 / 三焦经23个；肺经11个 / 大肠经20个。总数与一年天数相距不远，但也有说法，说穴位数目达到720个）。

躯体(前/背): 任脉/督脉

下肢(内/外): 肾经/膀胱经 肝经/胆经 脾经/胃经

上肢(内/外): **心经/小肠经** 心包经/三焦经 肺经/大肠经

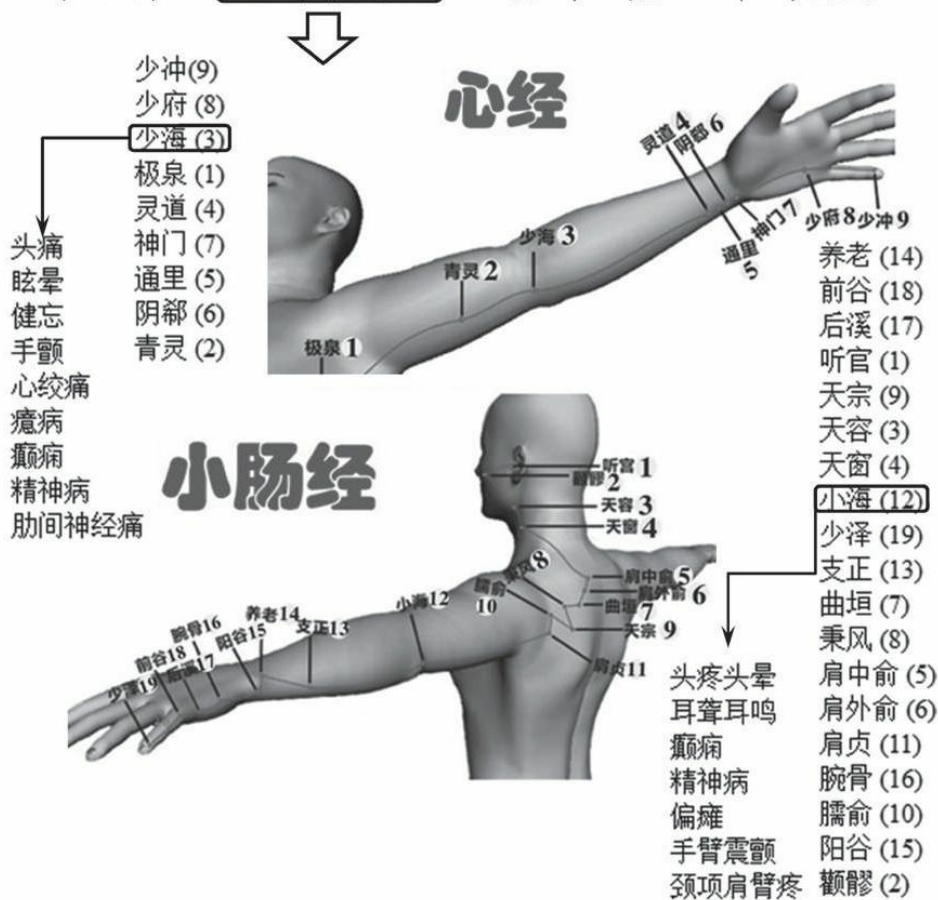


图1.47 人体上7组经络

按中医理论，每一个穴位对某一个症状或多个症状可能有缓解或治疗作用。一个症状依据其原因不同，可以与一个或多个穴位有关。足底反射区也类似。有某种症状时，依据严重程度，刺激对应的穴位，有酸胀、酸痛、刺疼等感觉，被刺激后可能出现临时塌陷和临时红肿等。如果症状变成了慢性的，对应的穴位刺激时，可能有气室感、颗粒感、条带感或肿块。

(5) 7种人际关系以及人际关系7原则

美国人事管理协会提出的人际关系，属于人与人在相互交往过程中所形成的心理关系，包括：①亲属关系、②朋友关系、③同学关系、④师生关系、⑤雇佣关系、⑥同事或战友关系、⑦领导与被领导关系。有时两人之间拥有多种关系，比如说同学也可以是朋友。

建立人际关系需要遵循的原则是：相互原则、交换原则、自我保护原则、平等原则、相容原则、信用原则和理解原则。

（6）常见的7种能量形式

我们的日常活动离不开能源，比如说照明、电脑和空调用的是电能，煤气炉用的是化学能，暖气片散发的是热能，太阳照射提供光能，风力发电用的是风的机械能。我们常见的能量形式为机械能、热能、电能、磁场能、光能、化学能和（原子）核能7种。能量的基本单位是焦耳和卡。

这里单单提一下机械能。机械能包括动能与势能，存在于运动物体、风和潮汐之中。动能等于质量乘以速度平方除以2。风速为1米/秒时，每立方米的地面空气携带的动能为0.6125焦耳。取北京市东西各100千米的长度，再取10千米的高度。那么总体积为 10^{14} 立方米。假设密度沿高度方向的变化不考虑，那么这个体积蕴含的动能为 0.6125×10^{14} 焦耳。一度电是一千瓦小时，即 3.6×10^6 焦耳。因此，这个风场蕴含的动能，对应 17×10^6 度电，即一千七百万度电。风力发电正是利用风的机械能。

（7）小世界理论与名人阻尼现象

认识一个陌生人只需要大致关联包括自己在内的7个人。哈佛大学

心理学教授斯坦利·米尔格兰姆设计了一个实验：米尔格兰姆在信中写有一个股票经纪人的名字，把信随机发送给各城市的一部分居民，信中要求每名收信人把这封信寄给自己认为比较接近这名股票经纪人的朋友。每位收到信的朋友将信寄给他认为更接近这名股票经纪人的朋友。结果表明，大部分信件都寄到了这名股票经纪人手中，每封信到达时平均经手了6.2次。

米尔格兰姆依据试验总结出了所谓的“六度分割理论”或“小世界理论”。依据该理论，如果你想认识一个陌生人，平均而言，只需要通过5个中间人就可以达到目的。两个需要相识的人加上5个中间人，一共是7个。这一理论对建立社交圈子和开展商务活动具有非常重要的价值。这只是早期在美国做实验的结果。

实际应用中，反映的是平均数或者普通人之间的概率。对于具体情况，不一定需要7个，也有可能永远达不到目的。例如，你不可能就单凭这个理论来认识某国领袖，或者电影明星。这里面存在一种所谓的信息传递阻尼现象。越有名的人，存在的阻尼越大。普通人不可能轻易去认识一个名人。虽然如此，六度分离理论中关联人的个数是7。

（8）算盘一档有7粒珠子

算盘每一档有7粒珠子，下面5粒，每一粒代表1，梁之上有2粒，每粒代表5。这7粒珠子能很方便地将结果在10以内的加减法完成。超过10时，就用相邻的左边一档代表1的表示10。用7颗珠子而不是10颗作为一档，既能完成运算又最省时省力，因此是一项伟大的发明，在推动古代文明进程中起了不可估量的作用。

文字中的数字7（图1.48、图1.49）

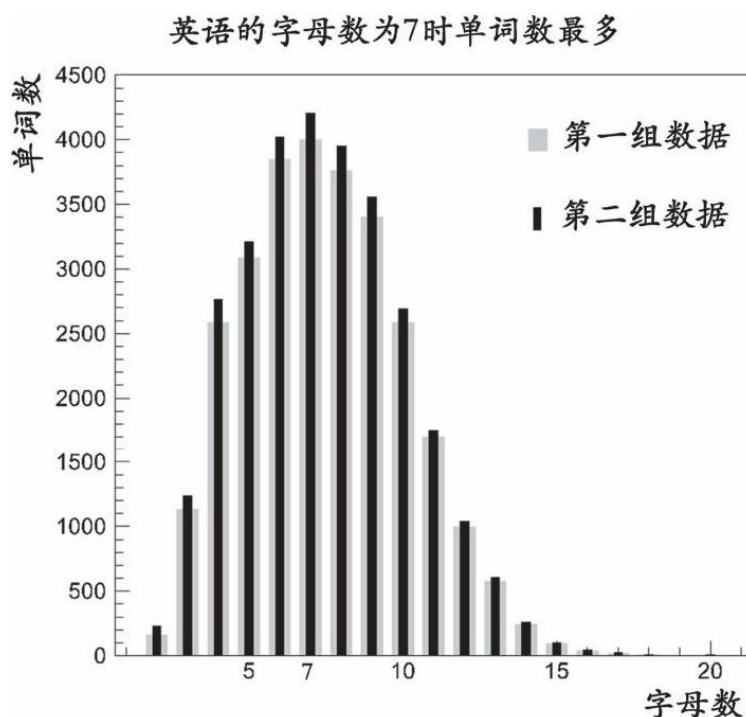


图1.48 具有1~21个英语字母的单词

有一份资料，统计了字母数从1~21的一些常用英文单词，有两组数据，一组比另一组更完整。对于两组数据，字母数为7的单词最多。

对于一篇具体文章，一些字母数较少的单词可能多次重复出现，因此一篇文章中单词的平均字母数一般小于7。杰姆斯·哈德利1858年那篇《数字七》（Number Seven）的英语文章有6839个单词，平均每个单词是4.73个字母。据说，更一般的统计表明，英语平均每个单词是5个字母左右。

期刊《文字改革》在1958年第3期发表了“常用汉字笔画统计”。结果表明两千个常用简化汉字的平均笔画大致是9.8画。中国第一首成熟的七言诗，第一句总共约55笔，平均每个字是8笔左右，第二句平均每

字达到11笔以上。

汉字的基本笔画数目、实词数目、虚词数目、句子成分数目、点号数目、标号数目等都在7左右，正负不超过2。

下面将要看到，人的短期记忆可能受限于7，正负不超过2。同样一篇文章，如果是英文，那么平均每个单词的字母数目一般小于7，而翻译成汉字后，平均每个字一般是7笔以上。因此，从这个角度看，英语更容易记忆好学。

因此，文字改革可以朝着减少平均笔画数目的方向发展，那样有利于记忆，学习的代价更小。问题是，让平均笔画小于7的同时，能造出那么多满足要求的汉字吗？利用现代图形学和人工智能技术，完全可以对汉字进行这样的改革。

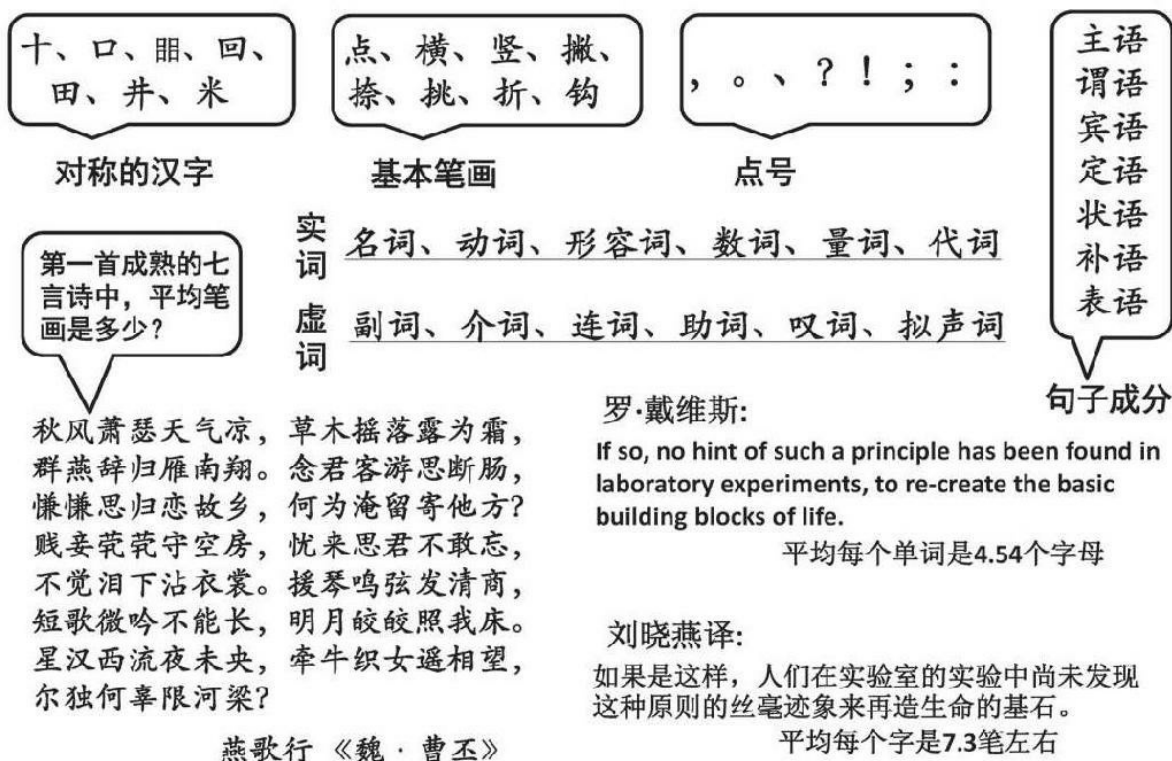


图1.49 汉字中接近或偏离数字7的现象

数字7的科学研究

不少人把数字7当作神秘的数字进行过研究。据说，12世纪葡萄牙人纳斯·玛尼德斯就解释了数字7是自然世界的神奇数字。历史上出现了关于7的一些重要文献。奥地利东方学者浦格斯塔（1848年发文）、耶鲁大学希腊语教授暨美国科学院院士哈德利（1858年发文）、俄罗斯神秘论学者布拉瓦茨基夫人（1880年发文）以及获得过美国国家科学奖的哈佛大学心理学教授米勒尔（1956年发文）等，均对神秘的数字7做过研究。

米勒尔创立了认知心理学这一领域，获得过美国国家科学奖。他用实验证明人的短期记忆能力受限于7，正负不超过2，发表的论文《神奇的数字： 7 ± 2 ；我们信息加工能力的局限》是心理学领域引用率最高的论文之一（到2015年5月22日为止，至少被7192篇学术期刊论文直接引用）。米勒尔法则可简单描述为：人在一次事件中，能记忆的对象（颜色、声音、气味等）个数平均为7。大家多数人一次能记住7样左右（而且通常是头尾记得较牢，中间的容易被忘掉）。因此，人的短期记忆能力是受限的，一次只能接受、处理和记住7条信息，误差2左右。研究结论被电信公司使用：电话号码太短满足不了要求，太长则记不住，于是设为7左右（区号除外）。

杰姆斯·哈德利（James Hadley）是耶鲁学院希腊语教授，美国东方学会主席，美国哲学学会副主席，美国科学院院士。在1858年发表的《数字七》的论文中，他试图从数字运算的角度解释7的特殊性。例如，数字7不能用其他两个数字相乘得到。在1到10的整数中，一些数字

相乘得到另外的数字（如 $2 \times 3 = 6$ ），或由其他数字相乘得到（如 $8 = 2 \times 4$ ， $9 = 3 \times 3$ ），但除了与1相乘得到自己，7既不能乘出10以内的其他数字，也不能由其他数字相乘而来。因此，7似乎是孤单的数字。有关数字7的一些奇特数学特性还远不止这些（不信，你用1除以7看看得到什么）。

不仅算术如此，几何上也有点奇特。例如，在一个圆外接半径与这个圆相等的圆，只能接6个，加起来正好是7个。正六边形及以下的正多边形，可以用圆规和直尺画出，但正7边形则画不出来。

数字7具有普适性的可能原因（图1.50）

记忆受限于7（正负不超过2）以及数字7的孤独性，可能预示人们在确定数目时，会在少了不够、多了记不住之中博弈，选择7可能是折中的结果。如果是由掷一对各有六个面、数字在1~6之间的色子（骰子）决定，则数字相加出现7的可能性最高。

物理学只有7个基本单位可能只是一种巧合。有关数字7的神秘性吸引了许多人的思考，人们希望找到更深层次的原因。然而，即使是发表了被高频次引用论文的米勒尔，他在论文《神奇的数字： 7 ± 2 ；我们信息加工能力的局限》下结论时依然很谨慎。他如此写道：

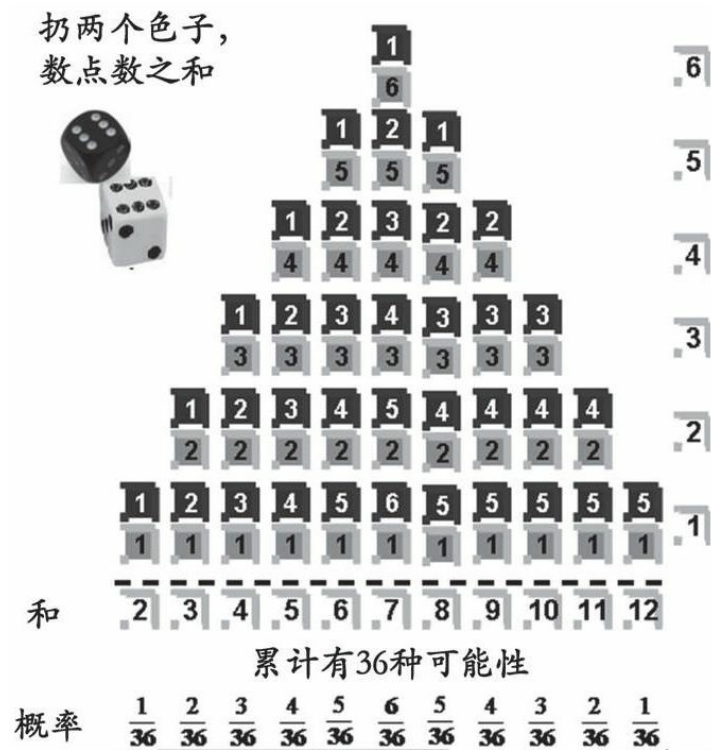


图1.50 扔骰子和为7的概率最大

最后，魔幻数字7有什么秘密？……，为何将人的年龄划分为7个阶段？为何是7个基本颜色，音乐是7个音阶？一周7天？为何是7点评价尺度？为何绝对判断可以分为7类？为何注意力覆盖7个对象？为何短期记忆跨度是7个数字？在这里，我建议暂不下结论。也许，在所有这些7的背后有更深层更深奥的原因，这个原因召唤我们去发现。但是，我怀疑，这也许只是一种没有价值的、毕达哥拉斯似的巧合。

米勒尔的这个论文，被后来的研究者写论文时，引用了近10000次。这些后来者，从本质上揭示了数字7的奥秘吗？吉勒斯·日尼亚克采用了85年之中的语言短期记忆数据（覆盖了两组测试数据，每一组近7000个对象）进行分析，结果表明，语言短期记忆能力平均值是6.56，正负误差为2.39。这与米勒尔法则给出的十分接近。

1.4 速度的限制 谁能把我们带往宇宙深处

运动和演化的快慢用速度表示，也用完成的时间表示。我们可以轻易到达遥远的星球吗？不能，因为速度有限制。不要一味地对这种速度有限制表示失望。正是速度的限制，使我们的环境较为安全。试想，如果我们的太空或地面，充满了极高速的物体，我们还没看见就飞来了，那多不安全。我们对时间的感受好像与真实时间有差异：排队时我们觉得时间极慢，而有好事的情况下，时间一晃而过。速度则真能改变时钟快慢。不信，你就乘坐接近光速的宇宙飞船去外太空旅行几年，回来时一定比留在地球的同胞妹妹年轻好多岁。

1. 移动了几倍体长或音速 尺度比例效应

在地面移动和在大气中飞行，你有多大，阻力似乎就有多大。于是，越大不一定越快，大象就跑不过兔子。即使大动物的速度表面上更快，也得看用什么方式去比。用人的奔跑速度直接与蚂蚁的速度去比，显得有点以大欺小。如果公平地比谁移动了多少个体长，结论可能反过来。万物速度可能不一样，现在给一个公平的舞台来竞速：比谁在单位时间移动的体长更多。当然，飞机可能不屑与蚊子相比，那就与音速相比。音速也就是小得看不见的分子玩接力棒的速度。如果物体快得连音速都瞧不起，那就与光速相比。光线中的光子其实小得没有大小甚至没有质量，但又有谁能和光子比速度？比完了，就会知道不要以大欺小，小的东西说不定更快，反应时间短，机动能力强。如果比转了多少圈，那就更显大巫是小巫了。

速度是几倍体长

如果我们认为重量比波音747客机小5000亿倍的果蝇的飞行速度比起波音每小时数百千米来简直是小巫见大巫，那就看看谁在单位时间内移动的体长更多。

成年果蝇的体长为3~4毫米，它们在24小时内会移动6.5英里的距离，大致相当于10.5千米，即1秒内移动120毫米左右，是30~40个身体的距离。长度为70.6米的波音747的飞行速度大致为每小时910千米，即每秒253米左右，1秒内移动的体长为3.58。因此，以每秒移动的体长计算，果蝇比波音飞机快10倍的量级。

尖尾雨燕飞行的速度最快可达每小时352.5千米，即每秒100米左右。其体长为9~23厘米。如果以23厘米计算，那么每秒最快可移动435个体长。

我们驾驶小汽车，在高速公路上限速110千米/小时，即每秒30米左右。以4.5米长的汽车为例，每秒移动不到7个车长。

非洲猎豹则是陆地上以绝对速度衡量最快的动物，每小时可奔跑120千米左右。但以相对于体长的速度，这不是最快的。澳大利亚一种虎甲最快可移动171个体长。加州的一种小虫（*Paratarsotomus macropalpis*），体长不到0.1厘米，每小时爬行半英里，每秒能移动322个体长。这被认为是陆地上具有最快相对速度的动物。

如果以这种相对速度计算，人的速度小得可怜，飞人博尔特1秒也就跑了6个体长左右。当然，我们完全可以用另外一种计算方法。如果我们比的是奔跑方向的身体长度，那只有10厘米左右（假设体型足够纤

细），那么博尔特那样每秒10米左右的速度（实际上他最快可以用9.58秒跑完100米），每秒移动了100个左右的身体长度（即身体朝前方向的长度尺寸）。这比波音747还快了几十倍。

弓箭射出时，初始速度大概是每秒40~70米。以1米长度的弓箭为例，这相当于每秒移动了40~70个箭长。一些枪射出的子弹初速度达每秒几百米。例如，AK47子弹初始速度为每秒700米左右。子弹弹头长度如果是2.68厘米，那么初始时刻每秒移动了2.6万多个弹长。这是动物没法比的。

速度是几倍音速（也称声速） 马赫数

飞机与小昆虫竞速，以某种标准看，居然败下阵来，那就和空气分子送声音的速度去比吧。这一比，就比出事情来了。原来，空气也不是好欺负的。你敢超过我接力赛送声音的速度，就会给你制造麻烦，除非你削平你的机翼，磨尖你的头部。

电影《绝密飞行》中有一段飞机起飞时的视频，你能听到1马赫，2马赫，3马赫，4马赫。马赫数是飞行速度与音速的比值。4马赫表示飞机飞的速度是4倍音速。在地面空气中，音速大致是340米/秒，即每小时1224千米。20千米左右的高空，温度比地面低80度左右，音速大概只有290米/秒左右，即每小时1000千米左右。在这样的高度以6马赫飞行，一小时就飞6000千米左右。地球赤道周长是4万千米左右。因此，这么快的飞机，数小时就能到达全球任一点，从北京起飞，不到两小时就到达巴黎了！

马赫是一个人名。之所以叫马赫数，是为了纪念他的贡献。据说马

赫不相信原子，且爱因斯坦还蛮欣赏他的。马赫数大于1时，是超声速飞行，小于1时是亚声速飞行。接近1时称为跨声速飞行。

对于飞机，马赫数不仅仅是一个简单的相对比值。马赫数的大小将空气流动分成完全不同的类型，以致高马赫数的飞机与低马赫数的飞机的形状差别很大，不是简单把发动机的推力提高，就可以获得高马赫数飞行的飞机。

原来，飞行器等物体撞击空气，引起的气压变化，以压力波即一般以音速向四周扩散。这种扩散力图释放局部的气压变化，推送到其他地方。但这种推送是在流动的空气中推送，或者说在推送时还受到飞机的进一步追赶与引起的新的气压变化的干扰。于是飞机的速度会影响气压变化的释放快慢或效率。这种影响就会与马赫数有关。

尤其是飞机超过音速后，即做超声速飞行时，压力波的释放速度快不过飞机，一般使压力波会在飞机头部堆积，形成冲击波堵在前面，尤其产生巨大的阻力（简称波阻），让飞机付出巨大的代价才能正常飞行。为了减小这种堆积，超声速飞机的头部要更尖细，机翼更薄平。

谁的转速快

要问每秒转了多少圈，或者转一圈要多少时间，一比还真是吓了一跳。银河系是一个螺旋结构，天体绕银河系中心旋转。转一圈，要两亿多年。这也太久了。地球绕太阳转一圈，365天左右。地球自转一圈，24小时。台风眼转一圈，1小时量级。表的秒钟，转一圈，1分钟。飞机机翼发出的翼尖涡，转一圈数秒量级（与飞机大小和飞行速度有关）。

澡盆涡转一圈，1秒量级。

蝴蝶翅膀转半圈，0.1秒量级。最快的蚊子，翅膀转半圈，0.001秒量级。

至于分子、原子、电子和光子之类的，它们的自旋与我们理解的宏观物体自旋有差异，是一种不能进行确定定义的状态。有人估算，如果用某种宏观等价方式来衡量自旋带来的效果，电子自旋一圈只需要 10^{-32} 秒的量级。

从这些数据看，旋转系统的尺寸越大，转速就越慢（即转一圈的时间就越多）。也许，把系统的尺寸乘以转速，得到的数，就差异不大了。

比例尺度效应 相似参数帮助你省时省力地解决问题

有一个科学理论，称为尺度比例理论（scaling theory），用到动物的速度，是指动物体重越小，那么它们相对于自身体长的速度就越大。自转和振动也是如此，转动物体越小，转速一般越快，振动物体的尺寸越小，振动的频率就可能越高。小的东西动起来快，就是比例尺度效应的一种通俗表述。

其实，有更多问题可能涉及尺度比例效应。比如说，飞机的大小也有尺度效应。大型飞机在结构上会受到尺度效应支配。升力以及承力面积与机翼面积成正比，因此与线性尺寸的平方成正比。而等于升力的重量却与体积成正比，即随尺寸按立方放大。因此，单位面积所承受的力（应力）随尺寸增大而增大。飞机尺寸巨大后，这种应力就非常巨大，结构无法承受。因此要么改进结构设计方式，要么限定飞机的尺寸。飞

机太大的话，对气流的扰动会更严重地影响下游飞机的起降。因此，飞机不是做得越大越好。

比例尺度效应还存在于空气中的落体之中。大物体掉落快，小物体掉落慢，就是因为阻力与面积成正比，而重力与体积成正比，从而越小的物体越容易达到阻力与重力平衡的终端速度，或者说下落速度越小。冰雹与雨滴就能快速落地，小得多的雾滴就会悬在空中很难下来。

我们学习物理时，可能提到小球和大球谁先落地的问题。如果回答说大球先落地，那就算物理概念错了。其实，让大的铅球和小的雨滴同时从12千米高度落下来，还真的是铅球先落地，雨滴很久以后才落地。这是因为，越小的物体，受到的空气阻力相对于其重量而言越大。

飞行的速度与音速的比，运动速度与光速的比，也是衡量速度大小的比例尺度，也刻画一些比例尺度效应。

马赫数也是一个比例尺度。上面提到，马赫数高于1时，会产生冲击波，这在马赫数远低于1时是不会出现的。也就是说，飞行导致的气流流动的性质，与马赫数有关。

飞机的速度接近音速时，即马赫数接近1时，飞行速度与声音速度的影响旗鼓相当，互不相让，会使问题变得更复杂。这就是所谓的跨声速情况。自然界也是那样，如果有两种因素起作用，弱的听强的，由强的因素主导。蚂蚁听蚁后的，蜜蜂听蜂王的。事情就很简单。如果两种因素一样强，谁都不听谁的，最后使事情异常复杂。振动的弦的两个频率接近时，会产生难听的声音。

反映两种因素相对大小的比值会决定自然规律具有不同性质，马赫数只是一个例子。科学研究过程中往往会去寻找是否有这样的相对比

值。如果相对比值一样就能决定物理规律一样，那这种比值就称为相似参数。科学家往往围绕相似参数的大小探寻是否有与相似参数相关的规律。这样往往能给出可表述、可推理和可应用的结果，大幅减少研究的盲目性，最有效地将规律性结果付诸应用。

2. 太快也不安全 幸亏有音障、热障与光障

物体越快，越能在有限的时间到达目的地。追求速度的提高促进了现代技术的发展。但是，阻力与障碍会阻止你恣意加速。虽然如此，对速度的妨碍也让我们世界更安全。如果物体在大气层试图接近或超过音速飞行，空气就如临大敌、提高气压众志成城地堵在前面。如果速度想进一步提高，就会把物体搓热烧毁。当然，正是因为这种火烧博望坡似的火攻技术，才使从太空高速进入地球的流星被提前烧毁，否则我们一点也不安全。在没有大气的太空中，速度可以无限制地增加吧。比如说超过光速100倍，让我们一年就可以飞到100光年以外的星球。不幸的是，宇宙有安全法则，你想接近和超过光速时，会有一种无形的力量阻止你，如果你胆敢接近光速，就让你超重得比宇宙还重。

音障与热障（图1.51）

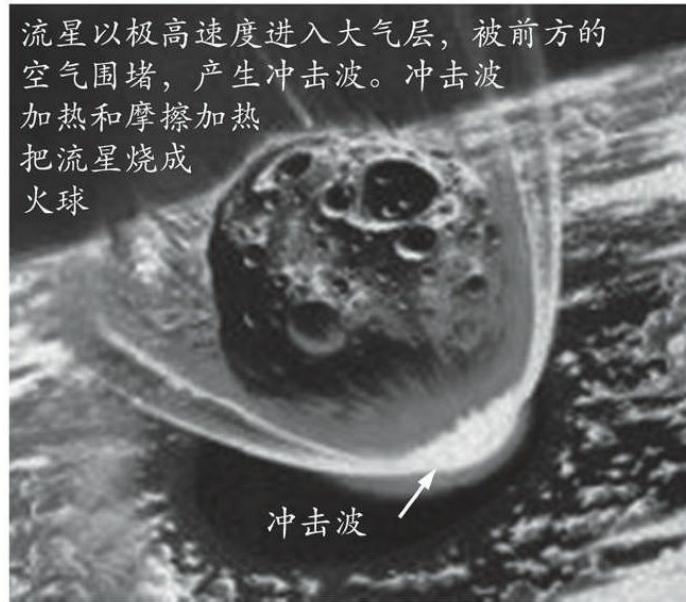


图1.51 高速物体穿越大气层会烧起来

物体在空气中运动，推动空气时会在迎风面增加气压。增加的气压一部分会以压力波即声波向四周传播。这种传播速度就是音速，在地面附近，大概是每秒340米左右。

超声速物体由于超过音速，前方被推动的空气无法通过那些传播速度等于音速的声波及时避让开，因此部分地被堵在那里，被堵的空气的前锋就是冲击波。冲击波像幽灵一样拦在那里，给物体施加额外的阻力，称为波阻。波阻与摩擦阻力合在一起形成巨大的总阻力，又因为早期发动机动力不够，使早期飞行器很难超过音速。这就是音障。

高速飞行的飞行器撞击气流，除了使气压增加，也使气温增加。除此之外，气流在机体上的摩擦，也会提高温度。我们使劲搓手掌，就会发热，道理是一样的。

由于是飞行速度与音速的相对大小决定了是否遇到音障和热障，因此就用它们的比值来表示速度大小。这个比值就是之前提到的马赫数。

飞行马赫数越高，飞行器表面气流温度增加越多。例如，太空物体以36马赫进入大气层时，头部温度最高点可以达到11000度左右。这么高的温度，当然容易将物体烧坏。天体中的流星以极高的速度落入地球大气层，产生极高的温度，使流星下落带着一团火球，绝大部分还没落下来，就烧光了。

因此，飞行速度接近和超过音速，遇到的阻力会大，以声音速度传播的压力波来不及释放，堵在那里设置障碍，这就是音障。飞行速度更高的话，摩擦就会将物体加热到可能将其烧坏的程度，使不加特殊考虑的飞行器很容易烧坏，这就是热障。

地球大气会爆炸和烧没吗？

高速物体在地球大气层中穿梭时，引起那么高的温度，空气会烧起来，但会爆炸吗？不会，否则一颗流星下来，我们早就炸没了。可是，流星不是带着一团火球吗？

极高马赫数的飞行器或流星体试图将空气加热到极高温度。被加热的氧气和氮气会用化学反应来吸收热量，而不是像煤气燃烧那样释放热量。

高温使分子热运动速度提高，空气中的氧分子和氮分子首先像弹簧一样振动起来，帮助消化掉一部分热量来降温。如果温度更高，那么分子热运动把双原子分子中由共享电子形成的化学键撞开，以便吸收掉更多的热量。

双原子分子被撞开成原子的过程称为离解。游离的氧原子和氮原子有了结合的机会，会产生一些一氧化氮分子，也吸收掉一些热量。如果

马赫数极高，使飞行器表面附近的气温升高到近万度，那么极高速度的分子热运动导致的碰撞，可以把原子中的电子打出来，形成带正电的氧离子和氮离子。这也会吸热帮助降温。这些离子与获得自由的电子形成的气体称为等离子体。另外，局部温度极高后，还会通过辐射散掉一些热量。

幸亏空气遇到高温产生的反应是吸热过程，才使我们的大气不会因为高速物体穿越而烧没。

光障

声速是每秒300多米的量级。如果我们觉得太快了，那就看看光速有多快。光速是每秒30万千米左右（精确地说是每秒299792458米左右）。从地面发出的光，1秒钟多一点点可以到达离地38万千米左右的月球表面。离地球1.5亿千米左右的太阳发出的光，8分钟左右可以到达地球。如果用地面音速那样的速度从地球飞到太阳那样遥远的地方，得14年左右。如果是更遥远的星球，比如说22光年以外的与地球有点类似的格利泽581d行星，必须得以接近光速的飞行速度才能在可接受的时间内到达。

高速飞行器遇到音障，是因为速度快了，导致空气用冲击波等设置障碍。茫茫的太空，除了偶尔遇到星际尘埃，几乎是什么都没有的真空。设想我们搭乘一艘宇宙飞船，每秒加速9.8米/秒，使我们感觉如同生活在地球上一样。差不多正好一年，我们的飞船就达到光速了。如果还嫌慢，那就每秒加速98米/秒，那1年下来，差不多10倍光速了。先不说用于加速的动力能否有这么大，能否持续这么久。自然界有一个常人无法理解的现象：任何有质量的物体接近光速时，会遇到光障。其

实光障这个名称并不常用，指世界上任何物质的速度不能超过光速，甚至无法接近光速。

我们之所以看到远处的物体，是因为物体发出的光朝你跑来。因此你看到的是须臾之前的物体。就像我们听34米以外的人说话，我们听到的是0.1秒之前的话，因为声音的速度大概是每秒340米。如果宇宙中有物体超过光速，那就乱套了。设想一个物体以超过光速朝你飞来，你还没看见它（因为他发出的让你看得见的光以光速向你传来），它就到了。因此，无法超越光速是一个宇宙安全法则。我们的宇宙不存在超光速的物质。

在加速器中，人们无论提供多大能量，都不能把比分子还小的质子加速到光速，更别说超过光速。电磁波、引力波以及光子以光速运动，因为它们没有静止质量。但没有超越光速的东西。

光速是宇宙中物体运动的最大速度，这个速度无法被超越。其次，也是最难理解的，光的速度相对于任何移动的观察者的，都是个不变的常数，在真空都是299792458米 / 秒（在非真空，比如说水中，稍微低一点点）。这种光速的（相对于以任意速度移动的观察者的）不变性，虽然是一种假设，但可以导出许多神奇的且能被验证的结论。比如说，一个立方体正对着你飞来，它变成了玫瑰花一样的形状。这些就是所谓的相对论效应导致的。

3. 相对论效应 卫星定位 核能利用

音障和热障，只是说困难，但技术的发展让人造飞行器得以突破音障和热障。光障却不能突破，不是技术问题，而是物理法则问题。不仅如此，物体接近光速和靠近有引力的星球，会发生一些匪夷所思的现

象。把地面生产的原子钟放在急速飞行的飞船上，原子钟走得慢了。朝月球急速飞去的飞船，看地球与月球之间的距离缩短了不少。这就是爱因斯坦提出的相对论效应。相对论效应是如此违背常识，以致人们说，地球上懂爱因斯坦的人站在一排，一个人一小时就能数完，也就是数千位的数目。依然有人还在反对相对论，虽然所有能做的实验都证明这个学说是正确的。连霍金都声称，他不断收到一些信，指出爱因斯坦错了。既然这么古怪，那就值得去看看到底会发生什么。除了加速器可以将非常小的粒子加速到接近光速，一般的物体的速度比光速小多了，因此拿去和光速相比，表面上显得毫无意义。因而，人们可能误认为我们的日常生活不会受相对论效应影响。事实并非如此，最著名的例子便是GPS定位系统，以及由相对论原理得到的质能关系在核能利用方面的应用。

速度让时间变慢、让距离缩短 狭义相对论效应（图1.52）

苏联宇航员谢尔盖当年以每小时28163.52千米（约7823米 / 秒，即1秒钟将近8千米的距离）的速度绕地球飞行了803天9小时39分。下来时，他的表居然慢了0.02秒。这不是表出了故障，而是由于所谓的物体相对我们以极快的速度运动时，上面的时钟会变慢的缘故。

一辆小汽车几米长，如果放在很远看，就显得很小了。但这只是视觉问题，小汽车本身还是那么大。但一个很快的物体从边上离去，该物体在运动方向的长度缩短了，这不是视觉问题，而是测量的长度真的变短了。

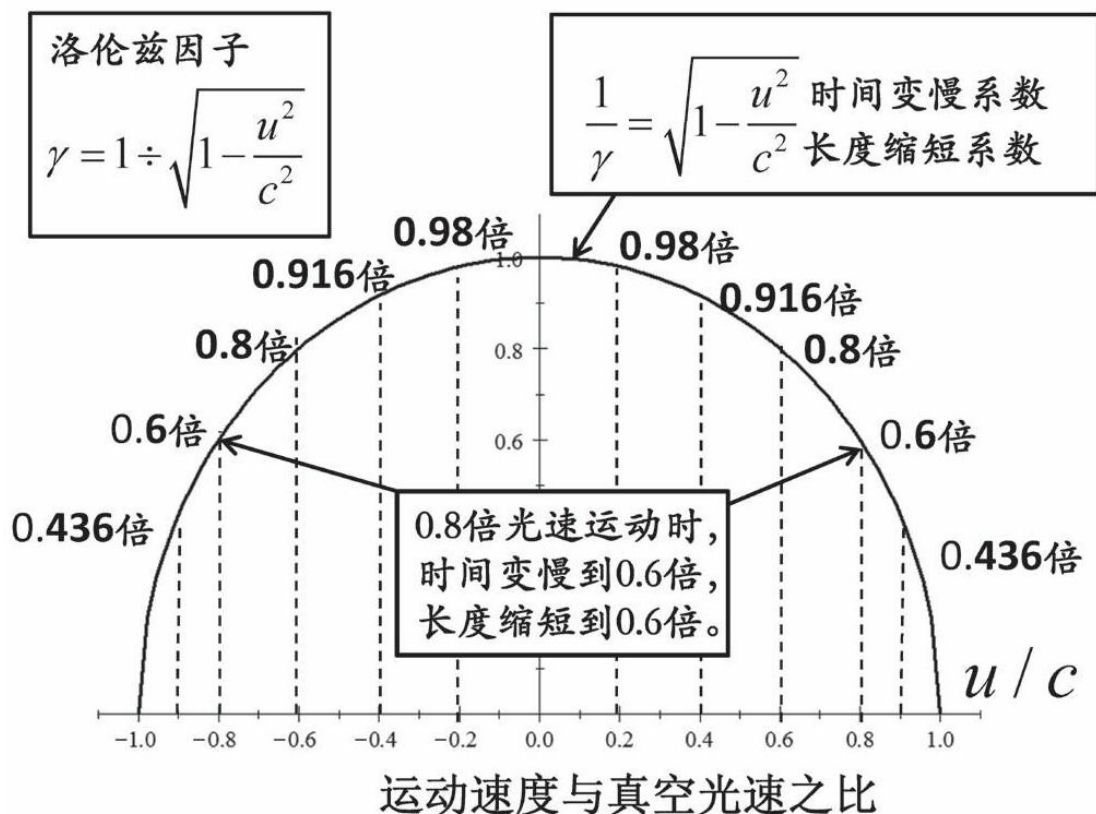


图1.52 洛伦兹因子与时间变慢与长度缩短系数

长度缩短了多少，时钟慢了多少，与运动速度相对于光速的比值有关，这是所谓的相对论效应。用1减去这个比值的平方，得到的数再开个根号（2的平方是4，3的平方是9；开根号就是把平方得到的数与平方之前的数对应起来，于是4开根号就是2，9开根号就是3）。再用1除刚才开根号得到的值，得到的数称为洛伦兹因子。例如，0.8倍光速对应的洛伦兹因子是1/0.6。洛伦兹因子的倒数（1/0.6的倒数就是0.6）就是长度缩短系数和时间变慢系数。例如，0.8倍光速的飞船上的时钟走过的时间只是地面时钟走过的时间的0.6倍。地球时钟走了1秒，飞船上的时钟才走0.6秒；飞船在飞行方向的长度，是起飞前的0.6倍。

纳闷的是，爱因斯坦是怎么得到这样的比值关系的。他主要用了物体的速度不能超过光速并且从任何移动物体上的观测者看运动物体（包

括光)的速度也不能超过光速这一假设。你在地面看一束光,它以光速离你而去。你开始追它,不管你追它的速度是多少,它相对于你的速度还是光速。

物体A以0.5倍光速相对于地球运动,物体A发出的光相对于A也以光速相对于A运动。但这束光相对于地球也是光速,而不是1.5倍光速。由于速度等于距离除以时间,爱因斯坦于是让距离和时间收缩,且收缩因子正好等于洛伦兹因子的倒数,就满足了速度怎么叠加都还是超过不了光速这一光速不变法则。

假设在地球和离地三十八万四千四百千米的月球之间有一个中继站,中继站离地球30万千米。速度为每秒30万千米的光(实际上也不是精确地等于每秒30万千米)从地球出发,1秒就到了中继站。现在假定飞船从地面出发,以每秒20万千米朝中继站飞去(忽略加速过程,也暂时不考虑能否达到这么快的速度)。

飞船出发时向中继站发出的一束光线,在飞船上的人看来,这束光是以每秒30万千米的速度向月球跑去。速度为每秒20万千米的飞船,发出的相对于飞船的、速度为每秒30万千米的这束光,我们地球上的人看,应该是每秒50万千米吧。不是,爱因斯坦假定,那束从飞船上发出的光,在地球人看来,速度还是每秒30万千米,1秒后到达中继站。

飞船朝中继站飞去,中继站和地球相对于飞船当然就朝相反的方向飞,即它们相对于飞船有了速度,速度也是每秒20万千米。飞船上的人测得地球和中继站的距离缩短了,不是我们看到的30万千米的距离,而只有22.36万千米。于是,它们发出的光线,以每秒30万千米的速度,只需要 $22.36 \div 30 = 0.745$ 秒左右就到达中继站。即飞船上的原子钟,只过了0.745秒左右,那束光就到达中继站了。这是飞船上记录的时间,是

原子钟的时间。

可是，地面上的人看飞船发出的那束光，还是每秒30万千米。地球上的人看中继站距地球的距离，是30万千米。于是，地球人看到那束光，是1秒钟到达中继站。

这就是说，地面上的原子钟走了1秒，飞船上的原子钟只走了0.745秒。这不是时钟出现了故障，而是相对论效应。这个效应说：运动物体上的时钟变慢了，且感知的距离缩短了。

这种运动的物体时钟变慢以及长度在运动方向缩短的效应就是爱因斯坦1905年发表的狭义相对论效应。在速度接近光速时，这种效应十分明显。

这个时间变慢是物理时间变慢，包括所有的生物、化学和物理过程，变慢了。距离也真的缩短了，在更短的时间可以跑完这个被缩短了的距离。飞船如果停下来，飞船上的时钟因此会滞后一些，但长度又恢复到原有长度。

引力使时钟变慢 广义相对论效应

由于引力使自由物体会获得速度，因此引力与速度有等价效应，也使时钟变慢，这就是广义相对论效应。引力越大时钟变慢越多。宇宙黑洞中心引力巨大，时钟就可能不走了。地面的时钟比高空的时钟受的引力大，因此高空时钟比地面时钟走得快。在速度较小以及引力较弱时，这种相对论效应不明显。

广义相对论效应是爱因斯坦1915年提出的。按广义相对论，引力也

使时钟变慢。可以这样理解，高处的物体在引力作用下会往低处加速下落，到了低处就获得了一个速度。这个速度的作用，就等效于狭义相对论运动物体的速度作用，使时钟变慢。因此，低处的物体相比于高处的物体，有一个等效相对速度，使低空的物体时钟相比于高空的，变慢了。

运动质量 质能关系 核能（图1.53）

爱因斯坦的学说还没完。运动物体的长度和时间不一样了，但并不表明运动物体上的物理化学规律不一样了。比如说，月亮上有人吃饭，你无论在地球上还是飞船上看，都是在吃饭。月亮上的物体满足万有引力定律，满足牛顿定律，满足热量从高温向低温传导的规律，无论是在地球上的人看来，还是在飞船上的人看来，都满足，且形式都一样。


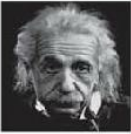
	创立者	时间与距离	运动定律	质量与能量
经典物理学	牛顿 	$T_{\text{动}} = T_{\text{静}}$ $L_{\text{动}} = L_{\text{静}}$	$F = m_{\text{静}} \frac{du}{dt}$ 牛顿定律	$m_{\text{动}} = m_{\text{静}}$ $E_{\text{动}} = \frac{1}{2} m_{\text{静}} u^2$
狭义相对论	爱因斯坦 	$T_{\text{动}} = T_{\text{静}} \div \gamma$ $L_{\text{动}} = L_{\text{静}} \div \gamma$ 相对论修正	$F = \frac{d(m_{\text{动}} u)}{dt}$ 相对论修正	$m_{\text{动}} = \gamma m_{\text{静}}$ $E_{\text{动}} = m_{\text{动}} c^2$ 质能关系
<div> T— 经历的时间 t— 时刻 m— 物体质量 L— 长度或距离 洛伦兹因子 E— 物体能量 u— 物体运动速度 $\gamma = \frac{1}{\sqrt{1-u^2/c^2}}$ c— 真空光速 </div>				

图1.53 爱因斯坦与狭义相对论、质能关系

可是，许多规律，包括牛顿力学定律（作用在物体上的力等于该物体的质量乘以该物体的速度随时间的变化率），用数学表述时，既包含位置（距离），也包含时间。比如说，在重力的作用下，下落物体获得速度，速度的大小就与时间和位置有关。既然运动物体上的时钟变慢了，长度在运动方向缩短了，那么将这些经典物理学规律用到运动物体上，就会出现与静止物体上的形式不一致的情况。爱因斯坦想，物理学与化学规律的形式不应该与物体（或观察者）是否运动有关。

于是，他把物理学规律加了修正，比如说把牛顿定律加了修正。修正后的规律，形式就与物体是否运动无关了，与观察者是否运动无关了。

修正的结果是，物体的质量必须与速度有关且运动物体的质量大于静止物体的质量。静止时有质量的物体，如果运动速度等于光速，那么运动质量就无穷大。光子的速度是光速，因此光子没有静止质量。

给自由物体施加作用力，物体会加速，速度越来越大，与速度平方成正比的动能越来越大。按经典物理学，这个动能的增加，等于力乘以力的作用距离。考虑了运动质量修正后，爱因斯坦发现，力乘以力的作用距离，累积起来等于物体加速后的质能减去加速前的质能。这里的质能是物体的运动质量乘以光速的平方，在运动速度较小时，运动质能与静止质能之差近似等于经典的动能表达式。爱因斯坦认为，这个加速前的质能就是静止物的能量，等于静止质量乘以光速的平方，这就是著名的质能关系。因此，静止物体有质量代表的能量，如果让质量消失，会释放巨大的能量。

达到光速时质量是无穷大，从而能量也是无穷大。因此你无法将一个有静止质量的物体，通过施加外力做功达到或超过光速，因为你无法得到无穷大的能源去推动静止物体达到光速。甚至哪怕是接近光速，都需要巨大的能量。目前没有发动机能做到这一点。

质能关系在核能利用的理论中能找到应用。原子核由带正电的质子和不带电的中子组成。强相互作用力的残余力将它们捆绑在一起，只有提供足够的能量做功，才能将原子核中的核子（质子和中子）拆开。这种能量就是结合能。于是将本来分散的质子和中子结合成原子核，就会释放能量。

一种原子中包含的核子（质子与中子统称为核子）的总数目称为原子量。原子量不同的原子，核子数目不同。核子数目不同，结合后各核子之间的排列方式与距离远近有差异。又由于强相互作用力的大小与距

离有关，因此不同排列方式导致不同核子之间距离不一样，从而结合能不一样。这就使结合能的大小并不正比于构成原子核的核子数目。

用原子核的结合能与原子核中核子数目相比得到的值称为比结合能。由于不同原子的原子核数目不同，排列方式不一样，结果是，比结合能的大小与原子量的大小既不是纯反比关系，也不是纯正比关系。原子量中等的原子核比结合能最大，拆开需要的能量就越大。这如同很强的人之间以及很弱的人之间不容易团结在一起，而中等能力的人更容易相互团结一样。

氢分子这样的原子量较小的原子核的比结合能较小，铁原子之类的中等原子量的原子比结合能最大，铀235和钚239之类的重原子的比结合能也较小。

质子和中子均有固定的质量。可是，一个原子核的质量，比构成它的那些质子和那些中子的质量之和要小，小掉的部分称为质量亏损。那么亏损掉的质量去哪里了？

原来，核子从分散状态结合到一起时，要释放能量，这个释放的能量就是刚才提到的结合能。释放掉的能量，与亏损的质量之间，满足爱因斯坦的质能关系。通过测量质量亏损，即可算出每一种原子的结合能大小。结合能以这种方式与质量亏损进行关联，直接验证了质能关系的正确性，也间接验证了狭义相对论的正确性。

核能的利用

如果能让铁原子之类的具有高比结合能的原子核分解或聚合成低原子量或高原子量的原子，那么显然需要添加能量做功，因为后者的比结

合能更小。这如同拆散一个具有高度凝聚力的团队或者合并几个这样的团队，需要付出较高代价一样。

反过来，如果让氢原子聚变成高原子量的原子，或让铀235和钚239裂变成低原子量的原子，就会释放能量。这种释放的能量就是核能。

太阳等恒星基于氢聚变释放巨大能量。由于氢的比结合能极低，因此聚变成高原子量时释放的结合能巨大，基于氢聚变的氢弹威力巨大。原子弹则是基于铀235或钚239的核裂变。原子弹涉及的裂变释放的结合能比氢聚变小不少，但威力也足够大。

核反应完成后，让剩余物质冷却下来，那么物质的质量减小量，按爱因斯坦的质能关系，与核反应释放的能量满足质能关系式。由于核反应释放的能量巨大，因此物质质量减小量是可观的，可以测量出来。相反，同等质量的物质因化学反应释放的热能，与核裂变相比，小1百万倍的量级，故质量减小量可以忽略。

恒星的质量非常巨大，热核反应产生的高温高压会阻止万有引力将拥有巨大空隙的原子挤压在一起。然而，如果恒星质量比太阳大许多倍，那么核反应结束后，在万有引力作用下，会坍塌成黑洞，将原子核挤压在一起，形成体积极小的致密星体。

GPS导航的相对论效应修正

离地心平均高度为26600千米处有24颗平均轨道速度为每秒4千米左右的卫星，作为GPS系统的星座。卫星上的原子钟，因相对论效应，每天会比地面同样准确的时钟快38微秒左右。如果不加修正的话，每天就会给导航带来10千米左右的伪距离。这对定位精度要求在数米以下的

GPS，显然不可接受。

卫星上的原子钟每天比地球上的快38微秒左右，不是来源于原子钟的故障与误差，而是因为速度带来的狭义相对论效应以及引力不同带来的广义相对论效应。这是一天的累计效应，每时每刻的具体效应还受离心力和旋转导致的相对位置或相对速度不一样的影响，后者称为萨格纳克效应。

虽然对于卫星这样的速度，与光速相比很小，从而变慢效应很弱，但每天积累起来，还是有几个微秒。同时，地面相对于地心也有自转造成的速度（赤道上为每秒0.46千米），也比地心时钟慢，但没有卫星慢得那么多。把卫星相对于地心变慢的时间减去地面相对于地心变慢的时间，就等于卫星时钟相对于地面时钟变慢的时间。按此理论，对于相对于地心以每秒4千米的速度运动的卫星，因狭义相对论效应，上面的时钟每天比地球表面的慢7微秒左右。原来，对于每秒4千米的卫星，时间缩短因子是 8.89×10^{-11} ，以此乘以1天的总秒数，是7.69微秒。

离地心2.6万多千米的卫星受到的地球引力，比离地心只有6千多千米的地球表面小10多倍，因为按照牛顿万有引力定律，引力与离开引力源质心的距离成反比。按广义相对论公式，受地球引力更大的地面原子钟，每天比在受地球引力小10多倍的卫星上的原子钟慢45微秒左右。

因此，因狭义相对论效应使卫星上的原子钟每天比地球上的慢7微秒左右，而因引力减小使卫星上的原子钟每天比地球上的快45微秒左右。合起来，卫星上的原子钟每天比地球表面上的快38微秒左右。

一架飞机可以从处在地球同一面的几颗卫星上接受不同时刻的卫星位置信息，并依据几何关系从几颗卫星的位置算出自己的位置。如果相

对论效应导致的卫星时钟时刻不加修正，飞机得到的就是错误时间的卫星位置，从而错误地算出自己的位置。不加修正的话，每天累计的伪距离会达到10千米左右，这与定位精度要求数米以内的要求相距太大。因此，人们会对相对论效应做出修正。一种做法是将卫星搭载的原子钟设计得走得更慢，每天慢38微秒左右，这样，相对论效应就使卫星上的时钟与地面时钟同步。

4. 时间很慢也很快

速度等于距离除以时间，因此也可以从时间看速度。时间在流逝，也有相对快慢。不仅是时钟决定了快慢，有时是感觉决定了时间的快慢。我们说，有时很慢，煎熬的时间慢如蜗牛。有时很快，快得如同时光流逝，以致大家会问时间都到哪里去了。这说明我们对时间的感觉与情绪有关。物理法则也会改变时间和生物钟的快慢，尤其是速度和引力可以将时钟调慢，让飞机和卫星上的原子钟走的时间和地面的不一样。弟弟坐接近光速的飞船去太空漫游，几十年回来后，会比留在地面的双胞胎姐姐年轻几十岁。

反应速度

发生在一个时间段的事件，不同人感觉到的时间长短不一样。动物都有自己的生物钟。事件发生的快慢用自己的生物钟测量，具有不同快慢的生物钟感知的快慢程度就不一样。如同一把尺子测量长度。如果尺子的度量长度很短，那么地球周长就显得很长，如果尺子的度量长度很长，甚至比地球周长还长，那么就会感觉地球很小。不同的动物，甚至不同年龄段的人，拥有的感知时间的生物钟快慢不一样，如同拥有的测

量长度的尺子不一样。

我们会觉得小动物很可怜，它们寿命和人类相比太短了。有的细菌生死就在一瞬间。如果生物钟的快慢与动物的体积成反比，那么也许不同大小的动物感受自己的寿命差不多都那么长，至少差异不是特别大。古人寿命短但生活节奏慢，虽然与古人相比我们现在的寿命按太阳年增加了，但由于现代人节奏快，一年时间显得很短，因此也许古人比今人更觉得一生很漫长。

反应时间也许是对生物钟快慢的某种有效衡量。正常情况下，人的反应时间为0.1~0.2秒。你听到号令枪响再起跑，这个时间间隔是0.1~0.2秒，这就是反应时间。据说，刘翔从听到号令枪响到起步，总的反应时间是0.139秒。司机的反应还包括对车作出操纵，反应时间是1.5秒左右。有的司机可能得3秒甚至更长。赛车手开车的反应时间肯定比1.5秒短多了。

动物和人的感觉器官接收信号（如眼睛看到或耳朵听到）或刺激后，将信息传递到神经系统，再由神经系统做出判断，向运动器官发出命令，最后由肌肉做出反应，这些环节累计的总的时间就是反应时间。各环节所需要的时间显然与参与这些环节的系统的尺寸有关，如同路程越短越快到达一样。

于是，反应速度似乎与动物的尺寸或体积成某种反比关系。大致而言，越大的动物反应速度应该越慢（反应时间越长），越小的动物反应速度越快（反应时间越短）。有时我们看到擂台上不同身高的搏击高手格斗，看不出身体越高的就一定占上风。也许身材矮小一点的因为反应快，弥补了打击距离和力度的不足。

体长1厘米左右的大齿猛蚁据说是地球上反应速度最快的动物。它们能在0.13毫秒内合嘴咬住动物，比人的反应时间快1000倍左右，比人类眨眼的速度快2300倍左右。

如果我们生物钟嘀嗒得极快，那么我们会感觉一天时间极其漫长。如果嘀嗒得飞快，那么连电影也看不成了，超声波也能听得见了。

原来，电影的画面按每秒24帧（有的可能不完全是这个数）一个接一个的输送，相邻两个画面很接近，但稍有差异。人眼及其数据接收与传输系统每秒能够输送10~12格画面，然后大脑的视觉处理中心会将每格画面保留1/15秒（视觉残留时间）。

当然，为了产生电影感觉，还需要在极短时间内识别每一副图像。在图像“刺激”视网膜后，大脑就会对形状、颜色和方向等信息进行处理。研究人员给受测试对象连续看6~12张各种类型的图片，如野餐图片或者带微笑的夫妻照片。每张显示时间只有13~80毫秒。结果发现，人的视觉能力比想象的要强，人脑能够解读眼睛仅看了13毫秒的图像，且眼睛每秒可以转移3次视线。研究人员还指出，大脑中的某些部位一定能将看到的信息保留至少13毫秒。

于是，视觉残留以及能瞬间解读图像，使我们感觉电影画面是连续变化的。如果我们视觉反应太快，那么我们会“看得清”一幅一幅不同的画面，就没有电影的动画感觉了（可能是皮影戏的感觉）。

人的反应速度快慢对于百米赛跑之类的运动非常重要。苍蝇的反应速度极快，因此能从人的眼皮底下逃生。

时间去哪里了

我们可能坐过飞机。飞机爬升的半小时显得很快，但飞机着陆前下降的半小时似乎显得很慢，显得过了很久。

小时候我们无忧无虑，很多事情我们还来不及体验，因此我们期待什么，却又不需要自己操心。这种等待会让儿童感觉一年过得很慢，寒假暑假总是显得姗姗来迟，过年过节久等不来。年龄大了后，我们承担的责任更多，需要靠自己来完成，因此总是感觉时间不够，久而久之我们就感觉到时间在飞逝。尤其我们老了以后，感觉所剩时间不多了，于是时间流逝更显快速。小孩度日如年，老人度年如日。

儿童时代身体活跃，生物钟较快，自己的生物钟嘀嗒60下可能比时钟嘀嗒60秒的时间短，于是感觉一分钟可能很长。老人生物钟慢，于是感觉一分钟很短。自己生活过的年龄也会成为我们的时间参照。一个三岁的儿童，才过了三年，一年与三年相比显得很长，因此儿童觉得一年很长。老人过了好几十岁，一年与几十年相比显得很短，因此感觉一年过得很快。除此之外，心情也可能临时影响生物钟。排队时、处理不愉快事情时，生物钟可能变快，时间显得很慢。遇到愉快的事情则反过来，生物钟变慢，时间显得很快。

心理学家通过研究得出了这些年龄越大感觉时间变得越快的结论，并且发现，这种感觉是跨文化的——在德国、奥地利、荷兰、日本和新西兰等国的实验中，参与者的表现十分接近。

一个已经过去的节日，在成人看来仿佛就在上周，而在儿童看来它已像是几年前的事。心理学家发现，如下几个因素影响到了我们对时间的感觉：①我们用难忘的事件衡量时间，年龄越大历经的难忘事件越多；②比例原理，即一年光阴占据生命中已过的时间比例越小，感觉一年过得越快；③生物钟随年龄增长而放缓，身体的衰老当然使跳动的生

物钟节奏更慢了；④年龄渐长琐事越多，使人们对时间的关注更少，因此从退休开始对时间快慢的感觉可能不会有大的变化了；⑤用做好一件事情的时间来衡量时间流逝，随着年龄的增大，压力增大以及琐事增多，让人感觉没有足够的时间做好一件事。

律

用标签找回时间（图1.54） 一万小时成才定

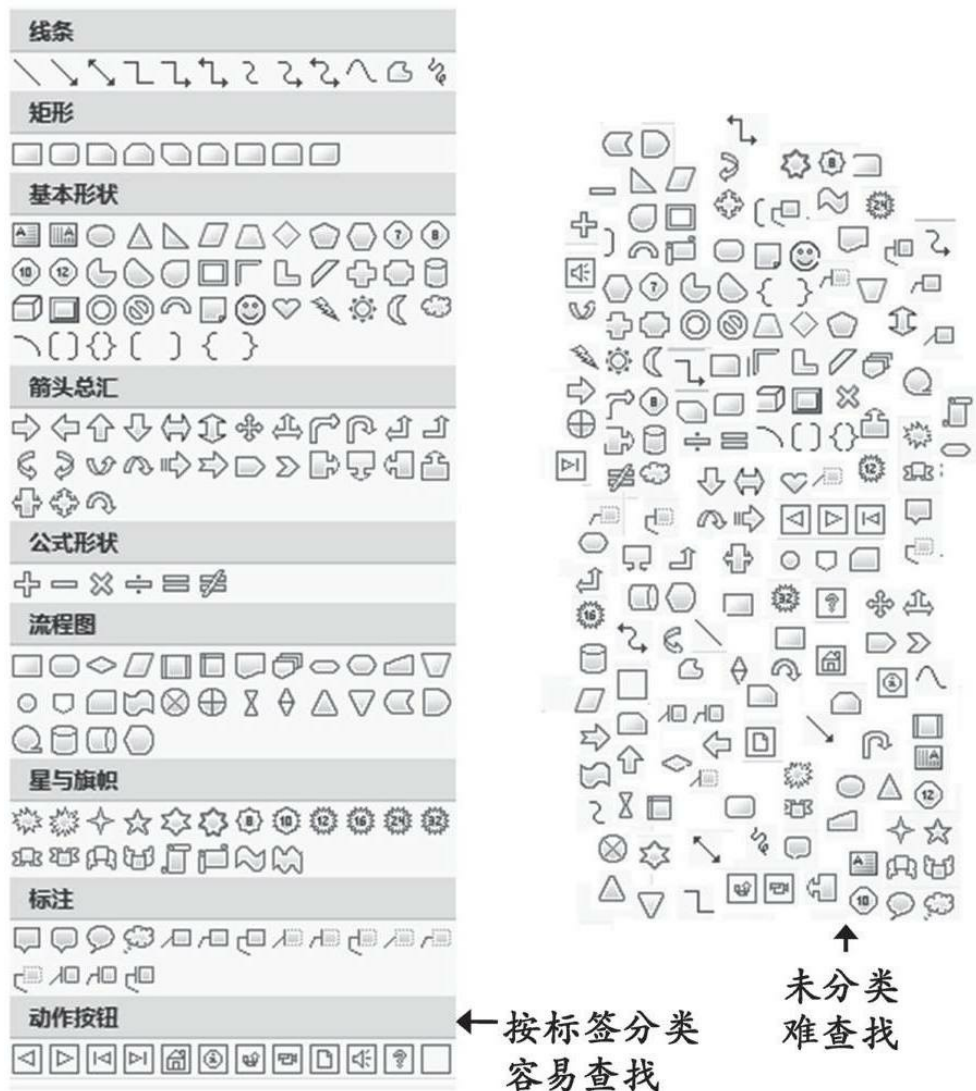


图1.54 Office中各种图形按标签分类布置

布赖恩·格林在科普片《扭曲的时间》中的开场白是这样的：人们说时间在飞逝，时间就是金钱，我们却在浪费时间，我们却在消磨时间；我们也试图节省时间，但我们对时间究竟有多少了解？他进一步说，时间就像河流那样，不断地朝着单一方向和朝着未来在流逝。

我们在问：时间去哪里了？我们觉得时间不够用，没法完成我们想完成的那么多事情。当然，我们也可能度日如年，如果在忍受令人不情愿的事情。我们能节省时间吗？当然能。

30年前，一位憨厚的大伯进了某校园，找他的侄儿。逢人便看便问，每分钟能遇到个把人，白天一天10多个小时能看到1000人左右。校园各类人，包括教职工、学生和其他人，也许是3万量级。照这种找人方法，理想情况下，需要30天左右才能把所有人筛选一遍，因此平均而言他需要15天才可能撞见他的侄儿。

30年前，一位女作家在北京旅游时要去校园找她表弟的女儿。首先她问清楚了表弟女儿在哪个系、哪个年级、哪个班。她上午11点左右进学校后，半小时内便打听到和找到了那个系，从该系教务打听到了表弟女儿的宿舍。那时正好是中午12点，她等在她宿舍那儿，很快她下课了。作家从进校园开始，一小时左右便找到了表弟的女儿。平均而言，比憨厚大伯节省了360倍的时间。

哪所大学、哪个系、哪个年级和哪个班，以及叫什么名字，就是给学生贴的标签。按标签分级筛选，可以将时间减小数个量级。找人只是一个例子，生活、工作、管理，尤其是探索与发明，按标签一级一级分类筛选，就能提高效率，节省时间。

有了可以支配的时间，就有可能成才。有一个异数定律，即一万小时成才定律，是指需要累计花上一万小时以上，才能学艺有成。博士研究生就读时间一般是4年以上，这个时间段足够提供一万小时让研究生成才。

每天也不需要一定达到8小时。十年以内，针对某具体活动每周练习20小时，每天3小时，就可以累计出一万小时。格拉德韦尔的研究显示，无论是学习古典音乐还是从事冰球运动，无论是作曲、打篮球、写小说、弹钢琴和下象棋，最低需要练习一万小时，才能成才。心理学家埃里克森调查了柏林音乐学院的学生，发现从5岁左右开始练习小提琴的学生，起初每周时间并不多，但随着年龄增长，优秀的学生练习的时间更长，到了20岁时，就累计1万小时，成为优秀的小提琴手。

飞机上的原子钟：海福乐和基廷实验

与人的感觉时间与实际时间之间有差别不同，狭义相对论效应和广义相对论效应真的会改变物理时钟，比如说，飞机搭乘原子钟，飞几圈后，走的时间会与地面的原子钟走的时间不一样。

1971年10月，海福乐（Hafele）和基廷（Keating）用飞机搭载原子钟做了实验。飞机向东向西做了两次飞行，两次飞行中，速度差别带来的狭义相对论效应有差别，而引力带来的广义相对论效应无差别，实验巧妙地同时捕获了狭义相对论影响和广义相对论影响。实验结果与爱因斯坦的理论预测非常接近。该项研究的理论预测与实验结果分两篇论文发表在1972年7月的《科学》杂志上。其中理论预测给出了很详细的计算方法。

飞机一次向东飞行，在总共65.4小时的时段内，累计飞行时间为42.2小时。由于地球自西向东自转，因此向东飞行时，飞机的对地速度加上地球自转带来的地面速度才能算作相对于地心的速度。按理论计算，由于飞上蓝天后引力较小，广义相对论效应使飞机上的时钟在实验时间内应比地面时钟累计快了144纳秒，而由于相对运动导致的狭义相对论效应，应比地面慢了184纳秒。这样，在飞机上旅行了的钟总共比地面的慢了40纳秒。这是按爱因斯坦狭义与广义相对论计算得到的理论结果。实际带上飞机的原子钟，比地面的慢了59纳秒。这已经非常接近理论结果，与理论值的偏差来源于一些理论近似和一些可能因素的忽略。

一次向西飞行，在总计80.3小时的时间中，累计飞行了48.6小时。广义相对论理论计算的引力减弱导致时钟变快179纳秒。但向西时，相对于地心的速度等于对地速度减去地球自转的速度，结果是飞机相对于可以看成惯性参照系的地心的速度比地面旋转速度要小。按狭义相对论，地面时钟比对地心速度更小的飞机上的时钟慢，于是飞机上的时钟累计比地面上的快了96纳秒。加上广义相对论效应，飞机上的时钟累计快了 $179+96=275$ 纳秒。这是理论计算的结果。而实验搭载的原子钟，比地面的快了273纳秒，与理论结果相当吻合。

旅行改变双胞胎姐弟的相对生物学年龄

无论是22光年远的格利泽581d，还是离地球1400光年的开普勒452，均是地球目前拥有的航天器无法到达的。22光年就是每秒30万千米左右的光，需要走22年才能到达的距离，大概在200万亿千米远的地方。以每秒16.7千米的宇宙速度飞行的飞船，需要飞行大致40万年才能

到达。因此，无法在有生之年飞到这么遥远的星球。

向远以光年计的遥远星球做星际飞行，宇宙飞船必须以接近光速的速度飞行，才能在可接受的时间内到达。海福乐和基廷实验是真实的实验。现在来一个虚拟的实验，也许未来真的能实现。这个虚拟的实验也称为双胞胎佯谬，有许多版本，但都是说明同样的问题：乘坐高速飞船的人，会比地球上的人老得慢些。

20岁的弟弟从地球出发以0.8倍光速去22光年以外的格利泽581d星球，20岁的双胞胎姐姐留在地球上等弟弟回来。假设弟弟加减速飞船以及滞留在格利泽星球需要的时间短得可以忽略。姐姐看远方星球是不动的，于是她看弟弟来回的距离是44光年。这个距离除以0.8倍光速，是55年。即弟弟回来时，姐姐过了55年了，已经75岁了。

弟弟在飞船上，那颗遥远的星球和地球是以0.8倍光速朝与他飞行的相反方向飞去。于是，距离缩短因子是0.6倍，在他看来星球与地球之间的距离只有 $22 \times 0.6 \approx 13.2$ 光年，来回距离约为26.4光年。这个距离除以0.8倍光速，就是33年。

于是，弟弟只过了33年，回到地球时53岁了。因此，弟弟回到地球时，弟弟为53岁，姐姐为75岁。弟弟这一旅行，比姐姐年轻了22岁。

时间旅行可能吗？

有时人们把运动物体上的时钟变慢解读为时间旅行。实际上，至少相对论中的时钟变慢，只意味着时间向前走的快慢之差，没有暗示时间可以倒流，不意味着可以回到过去。虽然如此，有些报道指出，2040年可能出现时间旅行机器。但也许只是一种媒体的不同解读，与能回到过

去的时间旅行可能不是一回事。

有几类思考指出，时间旅行不太可能存在。时间是所有人和所有世界的共有时间，怎么可能一个人或者一艘时间飞船拥有可以倒流的时间？如果未来技术的突破真的可以实现旅行，那么未来人早应该坐着时间旅行机器来造访我们了。其次，如果时间旅行真有可能，那么我们可以回到过去改变历史，这样会引起混乱和悖论。最著名的例子就是回到过去把自己杀死，既然被杀死了怎么又能活到未来再回去杀死自己。

人们说，从相对论满足的数学模型可以得到一些支持时间旅行的数学解。比如说链接两个时刻的虫洞就是广义相对论数学模型的一种解。一个虫洞可能链接今天和数千年前的古代，而穿越虫洞不需要多少时间，于是可以瞬间回到古代。

但有那样的解不一定有那样的事实。比如说，圆的面积等于圆周率乘以半径的平方。面积为3.14平方米的圆，对应的半径是1米左右。你取半径为-1米左右，代到圆面积的数学模型中，得到的面积也是3.14平方米。但没有一个圆的半径是-1米的。因此，只有具有物理意义的解才有意义。

也有一种说法，由于接近光速的飞船，其时间比地球上的过得慢，等于光速，那么时间就凝固了，因此外推到超光速的情况，那时间就倒流了。这种反推不一定正确。首先，超过光速本身不可能。其次，即使超过光速，物理学规律不一定是低于光速的物理学规律直接外推过去。另外，时间倒流是什么意思？是出发时身高1.8米且20岁的人，在飞船超过光速后，就回到1.2米的儿童时代？恣意外推，很容易得到一些古怪的结论。许多自然科学理论满足的数学模型其实也存在一些荒谬解（物理学家要排除荒谬解，而不能让数学家随意求解）。但时间旅行的

概念因为更容易引起大众兴趣，从而本身可能荒谬的结论也会被不断展望是真的并被寻求理论支持。试图支持荒谬结论的理论一旦高深到连反对者都无法理解时，这样的理论就可能红极一时。

但这并不是说，时间旅行就一定不可能。谁不期望能乘坐时间机器回到遥远的古代，欣赏战马奔腾的激烈场面？

5. 星际旅行 费米悖论

地球如果不独一无二，那么宇宙中一定有传说中潘多拉和坦尼斯那样美丽的星球。在不远的将来，最好在咱们的有生之年，甚至恨不得就在明天，咱们能探索遥远的宇宙天体，接触地球一样的系外行星，欣赏无与伦比的地外文明。然而，天体之间的距离之大，显得那么遥不可及，相对论揭示的宇宙安全法也阻碍咱们获得能在有生之年能到达这种遥远星球的速度。难道咱们就只能困在太阳系内？冲出银河系如果不行，难道不可以造访那些近邻，如22光年远的格利泽581d，离地1400光年的开普勒452？可是，千万注意啦，近光速飞行时，世界看起来变形变色了。前方的物体或背景会变肥变凶了，闪着可怕的蓝光呢。别怕，这只是相对论效应引起的光幻觉。物体远离咱们时，又会变瘦变温柔，发着慈祥的红光的。

近距离放大镜效应 运动物体颜色的多普勒效应 探照灯效应

先来了解几种常见的视觉效应。首先是近距离放大镜效应。一辆小汽车在几千米以外的地方，看起来像玩具一样那么小。到了眼前，它就像咱们人体还大。因此，在视觉中物体的大小，与离咱们的距离有关：

远而小近而大。

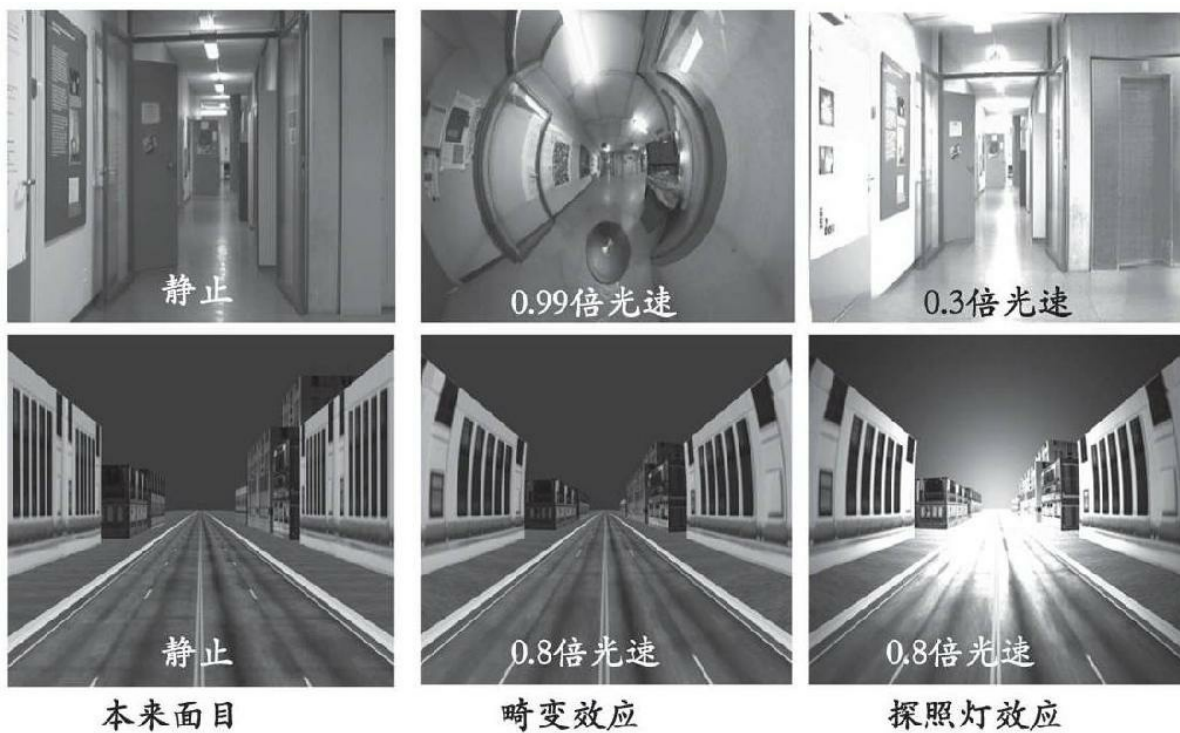
物体的颜色与物体反射或者发射的光的波长或频率有关，频率是波长的倒数。频率越高（波长越短）看起来越是蓝色、紫色，越低越是红色。当物体高速飞来，光波就更快地过来，因此频率显得越高了，所以是蓝色、紫色，远离时正好反过来，因此是红色。

探照灯效应或汽车前灯效应，是指由于这些灯太亮，使高亮度区以外的地方（包括汽车本身）看上去显得更黑了，于是耀眼的部分看上去更像白光。名人也具有探照灯效应或者汽车前灯效应。璀璨夺目让其他人显得无足轻重，一灯独亮虽然耀眼但也苍白。

这些效应在相对论距离缩短等效效应配合下，使近光速旅行时，看到的世界变形变色了。向你驶近的物体、街景的形状和颜色在你看来，与静止或低速时的不一样了，像哈哈镜中看人脸一样，物体显得变胖了，而且变凶了（发着蓝光紫光），远离时显得瘦了变温柔了（发着红光）。

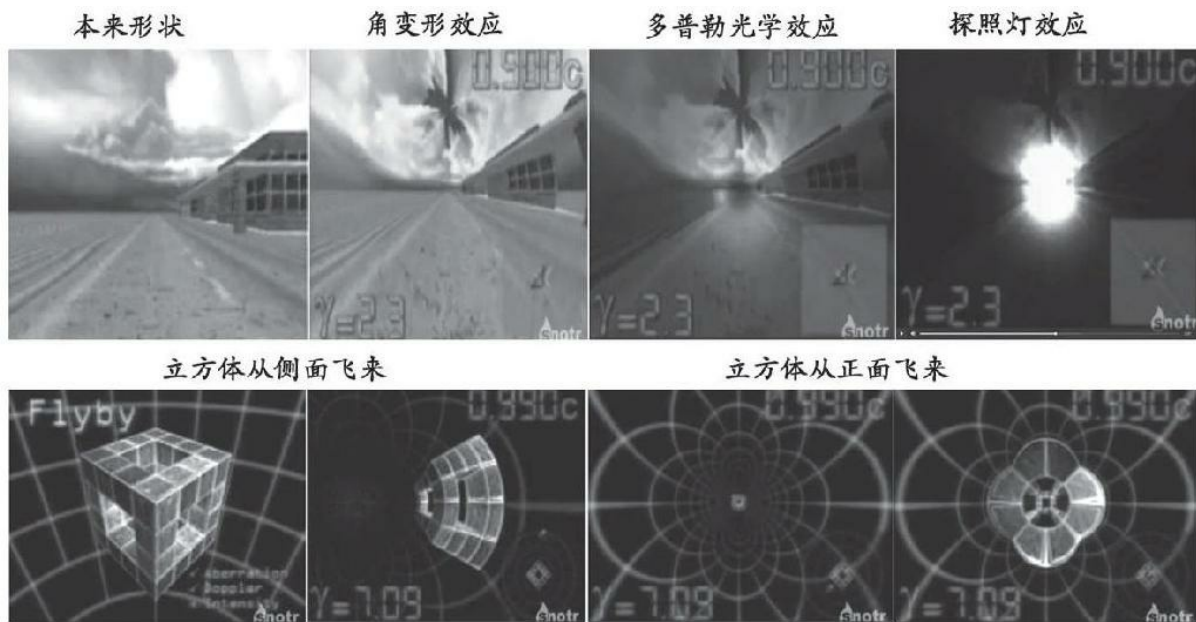
近光速旅行 奇形怪状和光怪陆离的世界（图1.55、图1.56）

扑面而来的物体，其不同位置相对于你眼睛的速度不一样。假设物体的中点对你眼睛的速度最大。物体上离中点越远的位置，对你的速度越慢。由于狭义相对论长度和距离缩短效应，相对你速度最快的点，相对论效应缩短的距离最大，因此显得离你最近，看得更大些。相对你速度更慢些的位置，相对论效应越小，因此显得越远，看得越小些。



(来源: ACM Trans Graph 等, 作者魏斯科普夫、克罗斯、科布拉斯和鲁德尔)

图1.55 近光速运动时的变形与变色效应



(来自snotr视频网)

图1.56 近光速飞行中看街景和近光速飞行的立方体

比如，矩形门洞框的中点就比门框的角点对你的速度快，从而显得更近，于是看起来门框的中部变宽了，或者说门框变胖了。明明是矩形，看得像圆形了。

街道上的两侧楼房之间的空隙可以看成矩形空间，由于上面的效应，与你眼睛同高度的部分显得宽了。于是，如果你在街道中以近光速沿着道路直行，两侧本来直立的高楼大厦看起来，其楼顶向街中心一侧弯曲。低矮部分接近你的速度更大，相对论效应使其离你的距离更近，两侧楼之间的间距显得就很宽。而两侧楼房的高点部分因接近你的速度更慢，显得更远，于是看起来更靠近街中央。

总结而言，运动物体上对你方位不一样的点因相对你的速度不同，而引起的相对论修正强弱不同，导致了物体不同点的远近视觉有差异，被看到的物体有角变形，即有畸变（**aberration**）效应。

由于多普勒效应，你看到正前方变蓝了，因为正前方的物体或背景相对于你的速度最大，多普勒效应最强。相对论长度缩短效应进一步压缩波长，加强这种多普勒效应。相对论效应还可以加强正前方（速度最大的地方）的亮度。这种耀眼的光具有探照灯效应或汽车前灯效应，使高亮度区以外的地方看上去显得更黑了，于是耀眼的部分看上去更像白光。

德国理论物理学家魏斯科普夫、克罗斯、科布拉斯和鲁德尔模拟了各种速度下，房间与街景的变形与变色。上面说的那种变形（畸变）效应使矩形门洞变胖了，街道两侧的楼顶向街中心弯曲。除此之外，多普勒效应使颜色变为蓝色、紫色，探照灯效应使前方或本来有光的地方显得更亮（因此蓝色和紫色不容易被肉眼看到）。

如果一个立方体一面正对着你眼睛从远处以近光速飞来，那么效应类似。明明是立方体的正面，即正四边形，看起来像从每侧棱边中点凸出的玫瑰花瓣。甚至是里面的往外翻，因为越里面相对你的速度越大，相对论效应越强，显得越在眼前。如果立方体从前方往侧边飞来，类似效应使立方体像空心圆管的一扇。

如果是圆球飞来，则周边各点接受的相对论效应一样，因此还是圆球形状。

星际旅行与文明等级

一百多年前我们能飞上蓝天，缩短了地球上的距离。几十年前人类到达月球，阿姆斯特朗个人一小步，人类一大步。虽然飞往太阳系内的火星目前还是一种奢望，但我们早就展望飞往更遥远的宇宙天体，如同星球大战与星际迷航等科幻电影展示的那样。

飞往以光年计的遥远星球，只能寄希望于能接近光速的飞船。我们不敢想象是否能把自已分解为光子，以便瞬间到达宇宙中任何星球，在那里重新组合成我们的本来躯体。我们身体发热时难免会发射光子，因此属于我们身体的一些东西早就造访了遥远的宇宙。

飞船速度接近光速后，按爱因斯坦质能关系，质量变得越来越大，动能越来越高，到了光速飞行，就无穷大了。以0.9倍光速为例，按爱因斯坦质能关系，将10吨的宇宙飞船，加速到这样的速度，大约需要 3.2×10^{14} 度电。这比将飞船加速到第三宇宙速度（16.7千米 / 秒），多8亿倍左右的能量。

获得这么高的能量，需要发动机提供动力推动飞船不断加速。可是，现有技术无法提供这么巨大的能量。为此，俄罗斯天文学家卡尔达舍夫（Nikolai Kardashev）提出了一个基于能量利用的高技术文明分类方法。

宇宙中可能存在三种文明，分别称为卡尔达舍夫等级 I 型文明、II 型文明、III 型文明。

I 型文明能像地球目前文明一样利用地球能源和部分地利用太阳能。实际上，地球人目前还无法全部利用地球能源，于是谦虚地称地球目前是 0.7 个 I 型文明。

II 型文明能利用所在恒星系的全部能量。如果地球人类达到 II 型文明，那么就能全部地利用太阳的能量。戴森（Dyson）设想：要完全利用一颗恒星的能量，需要在以恒星为中心、直径上亿千米的球面上布置足够多的人造天体来完全接受恒星发光辐射的能量。这个球面也称为戴森球。如果遥远的星球中拥有 II 型文明，他们可能构建了戴森球。戴森球上的人造天体将遮挡恒星足够多的亮度，以致地球上可以观察到。天文学家依据发光星球出现过的亮度奇异变化，猜想“发现了”戴森球。但这不过是一个猜想。

III 型文明能利用所在星系的能量。

银河系中有几千亿颗恒星，如果每颗恒星都像我们的太阳一样至少有一颗行星有通信技术文明（指具有射电望远镜技术），那么银河系中就应该有几千亿个高技术文明。卡尔·萨根在《神秘的宇宙》中对银河系高技术文明数目的估计，却远比这少得多。按他的估计，银河系当今顶多有 10 个左右的高级文明。

外星生命或文明如果存在，必须处在宜居星球上。只有那些离恒星不远不近、温度适宜、和地球差不多大，从而在星球表面能有液态水的行星才能是宜居星球。行星“格利泽581d”距离地球约22光年，其环境似乎如地球般符合孕育生命条件。开普勒452的直径是地球直径的1.6倍，位于距离地球1400光年的天鹅座，绕恒星一年大约385天，与地球一年365天很接近，有可能拥有大气层和流动的水。

费米悖论 不明飞行物（UFO）（图1.57）

由于人类尚无法造出能利用星系能量的宇宙飞船，因此盼外星高等文明来造访地球。只有等级高的文明才有可能利用足够的能量将他们的宇宙飞船加速到接近光速的飞行速度，造访其他星球。

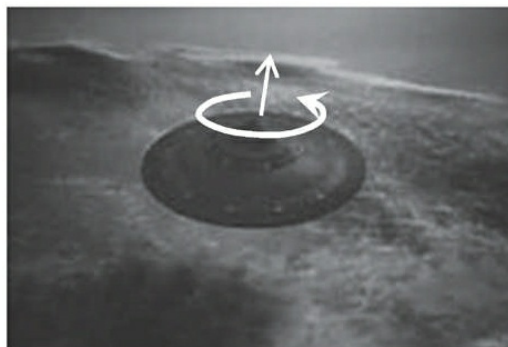


图1.57 旋转的飞碟

由于宇宙中有那么多的星系、恒星和行星，因此按理有相当多的行星满足孕育生命的条件并进化出了高等文明。有的行星出现生命的历史应该比地球的长多了，按理其文明比地球文明先进多了。这种高等文明的生物早应该造访地球了。可是，从来没有发现外星文明造访地球。这就是所谓的费米悖论。

虽然如此，许多人还是相信外星人乘坐不明飞行物造访过地球。我们许多人相信有外星人，但似乎谁都不太相信别人讲述自己遭遇了外星人或者不明飞行物的故事。

不明飞行物（UFO）本来是指不知道是什么的飞行物，但人们习惯地认为不明飞行物就是外星飞船。星星、极光、幽闪、流星、发光的鸟与昆虫、大气异常现象以及人造飞行器往往被摄像机捕捉到而误以为是UFO，但普通百姓倾向于相信甚至期盼真的有外星人驾驶UFO来造访过地球。

甚至有相当多的目睹外星人的报道。短小的四肢、硕大的眼睛、虫子一样的外表，是科幻电影以及民间报道描绘的外星人的形象。如果我们承认伴随技术的进步和生活水平的提高，地球人类正在变得越来越美丽，那么技术文明先进得多的造访地球的外星人就不应该长得丑陋不堪，他们的容貌难道不是惊如天人？

我们在科幻电影中看到的不明飞行物呈碟状，飞行时在旋转。旋转也许不仅仅是为了产生陀螺效应，使遇到扰动后更稳定。也许旋转是为了产生离心力，这种离心力的作用如同地球引力。人类在地球，习惯了地球引力的生长环境。到了太空，就只能靠加速或者旋转来产生这种等效引力了。

宇航员到了太空，失去了重力，长时间会导致骨密度下降，肌肉流失和视力下降等。除此之外，我们在地球表面处于重力环境下的一切生活方式都出现麻烦。例如，我们用杯子盛水，杯子要立着。在重力作用下，水就留在杯中。可是在失重的环境下，杯子朝哪个方向放，水都无法装在开口的杯子中。如果空间站是旋转的，那么产生的离心力就可以模拟重力环境。如果不明飞行物的出现真的是外星文明造访地球，那么由于地球重力环境与外星人所在的星球的不一样，他们就无法长时间出现在地球上，而只能像空间站一样留在绕地球的某种轨道上，通过飞碟旋转产生必要的离心力，模拟他们习惯的重力环境。当然，如果飞碟尺寸太小，那么高速旋转的飞碟也会使里面的人晕眩。

第二篇 柔情似水

江河湖泊与大海，让我们看到了丰富多彩的水面世界。那里拥有奇妙无比的涟漪、旋涡、水浪。水面生物与船只穿梭时，又会给水面带来波动，进一步丰富水面现象。水流冲刷如同大自然的雕刻师，让河道弯曲，让小湖出现，让不同形状的棒棒糖不可思议地演化成同样的形状。澡盆涡似乎谁都理解，唯有严谨的科学家不敢轻易地把它放在教科书中讨论。

水滴落在荷叶上，能保形成水珠。一只小小的水黾，靠与荷叶表面细微结构相似的小毛毛撑开水面，能在水上凌波微步，激发的涟漪形态居然与超声速飞行扯上关系。其求偶、威慑和驱赶利用了涟漪水位一高一低的交替变化的频率，与现代通信和计算机存储共享相似原理。

提到大海，我们会想到鳞次栉比的海蚀地貌、引人入胜的蓝洞、五彩缤纷的礁石，浪花与白云交相辉映。然而，无风不起浪，无风三尺浪，这种矛盾的表述并不是文学语言为了渲染而牺牲了科学的严谨，是真真实实的存在。

水面甚至水下有多少未知世界不得而知，已知的世界也只能在这里涉及点点滴滴，尤其是水面上能看到的现象。

2.1 柔情似水 凌波微步的小昆虫

水是液体。我们可以立在陆地上，在水面就会沉下去，于是我们对水有一种恐惧。但我们都知道，上善若水，水是生命之源。构成水的分子长着一对小胳膊，招手就是液体，搭手就是冰块。粘上荷叶就变成水珍珠，露在外面就筑膜给小昆虫提供蹦床，让水黾凌波微步。怪不得我们说柔情似水，想想杜甫诗中的点水蜻蜓款款飞。《传统文化论坛》提到老子之言：圆必旋、方必折、塞必止、掘必流。这就是说，水能善解人意，你提供什么舞台，水就演出什么花样。小昆虫戏水，水会羞涩地泛起涟漪。涟漪中波纹沉浮，就是一种通信的方式，让水黾能发出驱赶、威慑和求偶的指令。

1. 柔情似水

我们离不开水。正是因为有水才使地球是宜居星球，是水孕育了早期生命。陆地上的动物和植物需要水分才能生长，海水的保温作用使我们的冬天不会太冷。水被植物吸收，在阳光照射下引起光合作用，释放出我们呼吸需要的氧气，其中的氢原子与被植物从大气中吸收的二氧化碳结合成淀粉和葡萄糖，又成了我们食物的主要来源。水让古人能远航，促进了地球人类的早期交流与文明的扩散，也承载了现代大型货物的运输。水面的水分子在上面空气分子的退让下，筑起让小昆虫能在水面栖息的膜。水用涟漪、水浪、旋涡等谱写美丽的画卷，激发了艺术家的灵感，也让我们赏心悦目。冷了，她凝结成冰块，让冰块浮在上面抗寒。热了，她悄悄蒸发，飘向天空，化作云雨去滋润陆地上的花花草草，使江河川流不息。大风起处，怒涛不止，巨浪滔天，减弱了风的速度，让陆地上的生物免遭灭顶之灾。水压、浮力和表面张力，她浑身都

是力量，让船能浮在水面，让小虫子能爬行，让小水珠晶莹剔透，让游泳选手有了竞技的舞台。

挥手就是水，搭手就是冰（图2.1、图2.2）

将铁块扔到烧化了的铁水中，铁块就会沉下去。这说明固体一般都比同样物质的液体密度大。水却例外，它结成的冰比自己轻，因此冰块会浮在水上。一般的固体都很坚硬，但薄冰可能很脆。有一副古联：兵下冰，冰上兵，嘎吱一声兵压冰。是指士兵在冰上行走可能把冰踩烂。

水是一粒粒水分子组成的。1纳米是1米的十亿分之一，水分子的直径约为0.4纳米，可见水分子有多小。一个水分子是由一个氧原子和两个氢原子组成的。两个氢原子就像两个小球，用化学键连在比氢原子大得多的氧原子球的边上。可以近似说，氧原子和两个氢原子分别处在一个三角形的三个顶点上，其中氧原子所在的顶点的内角是104.5度左右。看上去就像一个人在伸懒腰或试图拥抱的样子。

氧原子带的正电比氢原子多，把更多的带负电的电子吸引到更大的氧原子一侧了，于是水分子带有正负电极。

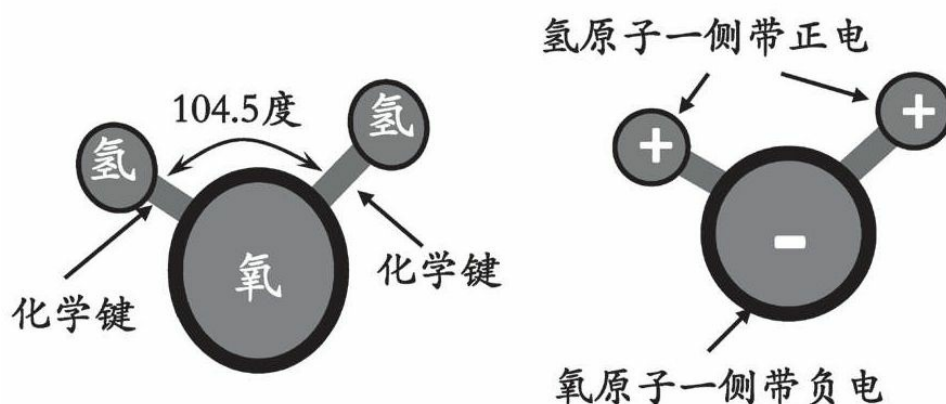


图2.1 水分子模型

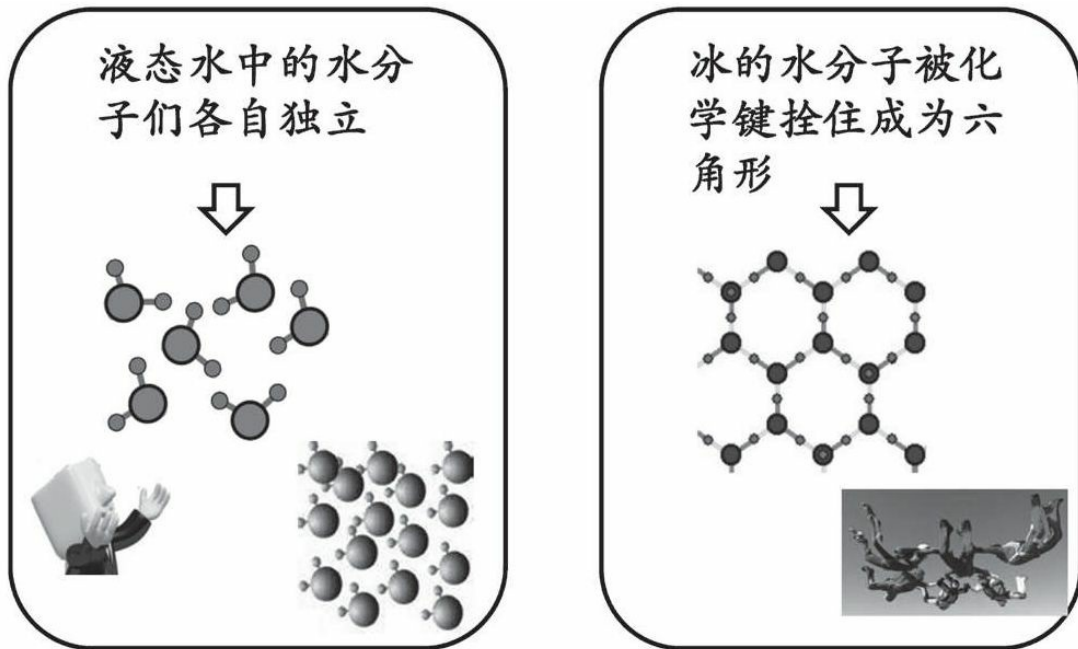


图2.2 液态水分子与冰分子

氧是由带正电的氧原子核与带负电的电子构成，氢是带正电的氢原子核与带负电的电子组成。水分子中的氧原子个头或者力量大一些即带正电的质子数目比氢原子多一些，把更多电子地拉到自己一侧，因此靠近氧原子那段带有一点负电，每个氢原子那段带有一点正电。这种正负形成的电极，在相邻水分子之间产生了静电吸引力，即内聚力，使水分子在一起，但分子之间却没有连在一起，因此形成可流动的液体。

如果水遇冷结成冰，那么就不是一个氧原子链接两个氢原子了，而是六个氧原子处在一个六角形的各个角点上。六角形上相邻两个氧原子之间有一个氢原子。多余的氢原子则与和该六角形不在一个平面的那些氧原子相连。在一个平面内，每个氧原子又是相邻三个六角形的公共角点。这样，一系列的六角形铺满平面。这样的图形最省料，也可以说，同样多的水分子情况下，冰这样的六角形结构占据的体积最大。于是，

水结成冰后，体积增大了，密度减小了。温度为零度时，冰的密度比水的小大概8.3%左右。冬天由于水底的保温作用，水下温度比水面上高，所以先从水面开始结冰。冰块由于比水轻，因此能浮在水面上。冰的温度越低其密度越高，但还是超不过水的密度。

由于将氢键拆开需要能量，因此冰化成水就需要吸收能量，反过来，水结成冰就会释放能量。

可以说，水遇冷后，用最省料的方式抱团取暖，并将余热奉献给外界，结冰时会散发热量。招手就是水，搭手就是冰，是指水分子中的两个氢原子像在招手，在冰中氢原子搭在两个氧原子之间。

分子巧手织绫罗 表面张力（图2.3、图2.4）

在水下，相邻的水分子之间由静电吸引形成内聚力，但各个方向的水分子吸引一个水分子，使作用在每个水分子的合力等于零。

可是，在水表面与空气结合的地方，水分子与空气分子之间的吸附力较弱，从而留有一些极性去增强与水面平行的水分子之间的内聚力。这种被增强的平行于水面的分子吸引力就是表面张力。形象地说，这相当于在水面上有了一层有张力或弹性的塑料膜。

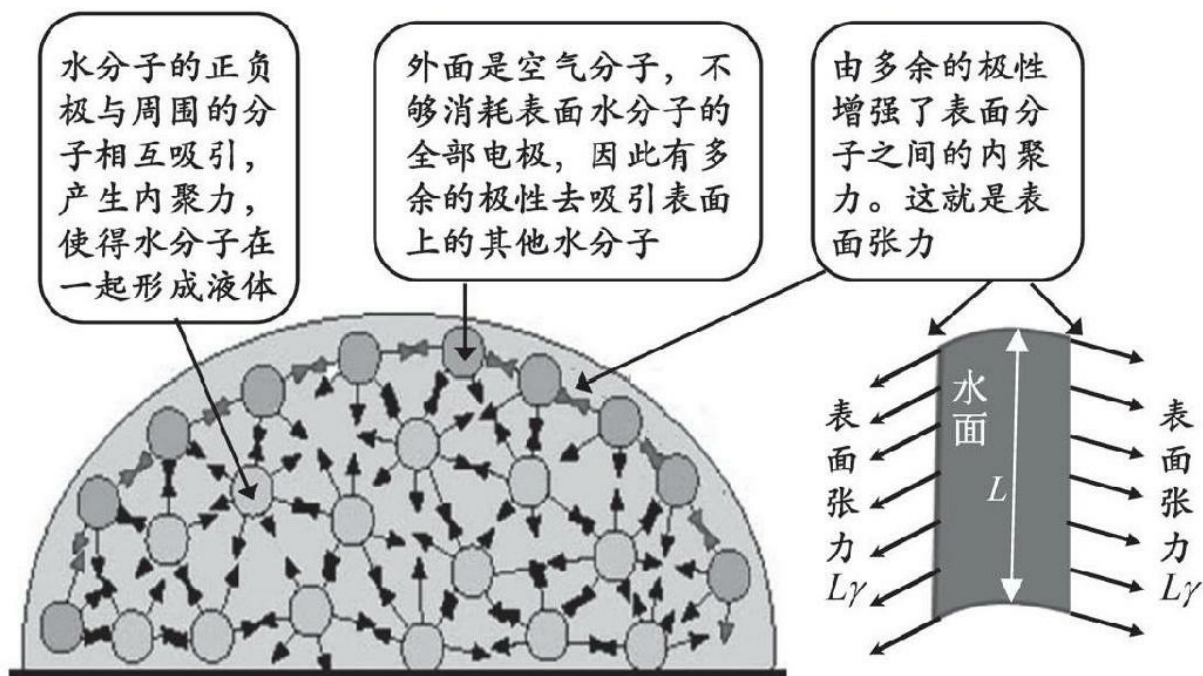


图2.3 表面张力来源于剩余极性

水面横截一刀, 刀口的表面张力就与刀口长度成正比, 也与表面张力系数成正比。暴露在空气中的水面, 温度为 20°C 时, 表面张力系数为 0.0728 牛顿 / 米。温度越高表面张力系数越小, 例如, 0°C 时为 0.0756 牛顿 / 米。长度一厘米左右的圆柱, 平放在水中, 两侧吃水线能产生的表面张力大概各是 0.75 毫牛。

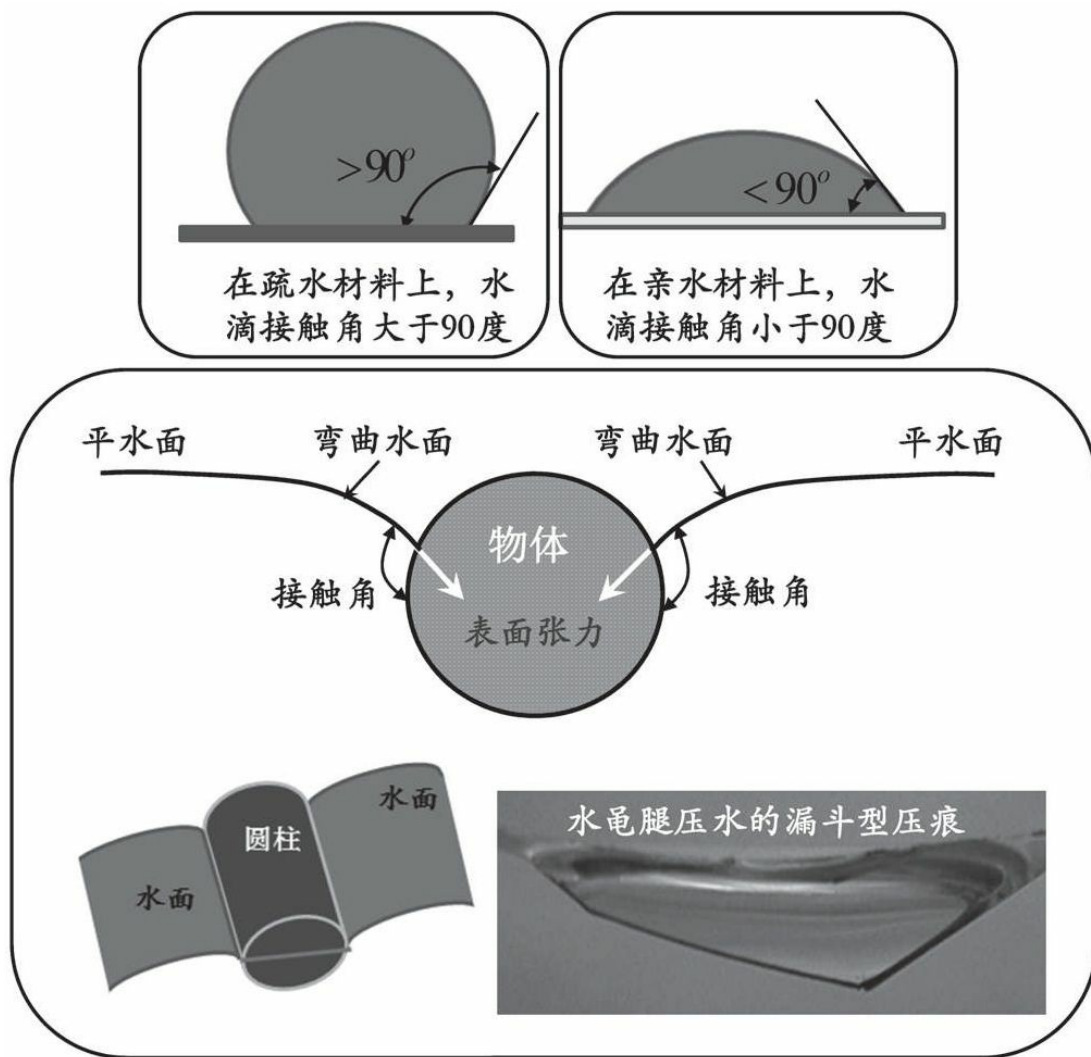


图2.4 疏水材料和亲水材料表面张力影响

如果水面弯曲，那么与水面处处平行的表面张力就试图将弯曲的水面拉平。于是，没有风的水面会平静如镜。如果是小水滴，表面张力的作用就如同小水滴被塑料膜捆住了一样，各点受力没有差别，因此使小水滴尽可能呈球形。

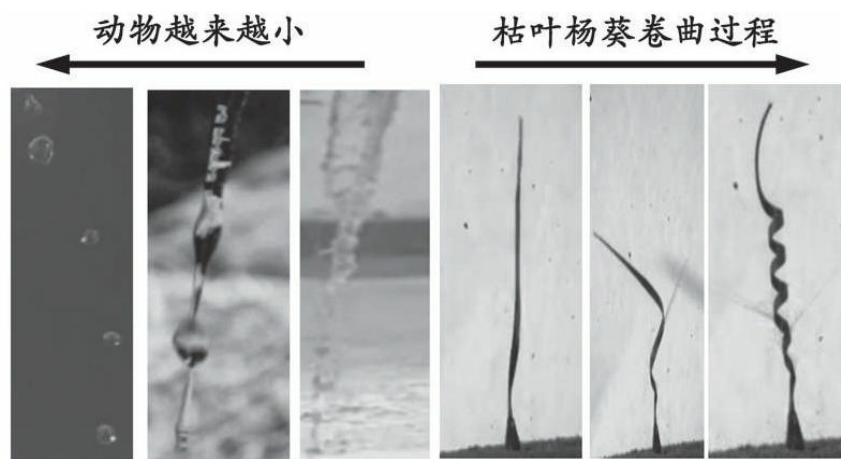
如果将小水滴放在固体材料表面上，那么具有易流性的水就会试图打湿表面。在干湿分界线上，固体的分子吸引力、水分子的吸引力就会同时发生作用。具体作用机制需要很深的表面化学知识才能解释。最后

的结果是，只有接触固体表面的水面与固体表面形成一个夹角（在水里侧的这个夹角称为接触角），才能维持平衡。如果接触角小于90度，那么水滴就铺展得很开，成了高度较小的半弧形，这种材料称为亲水材料。如果大于90度，那么如同小球被截掉了一小部分，这种材料称为疏水材料，也称为憎水材料，即不喜欢与水贴在一起的材料。

随处可见的表面张力影响（图2.5）

表面张力的影响随处可见。这里举两个例子。第一个例子是动物小便。小的动物排小便时，由于排出的尿液较细长，因此表面张力容易将其包裹出一粒粒小水珠。大的动物排小便时就不会这样。虽然如此，哺乳动物无论大小，排小便的时间都差不多。那位研究了蚊子战雨滴的科学家胡立德及其合作者发现，哺乳动物排小便的平均时间为21秒，围绕平均值的误差在13秒以内。排尿系统的体积在排便过程中，可以变化3600倍。

大的哺乳动物尿量大，但重力影响大，因此尿液流速大。小的哺乳动物由于排便管道小，虽然尿量少，但摩擦阻力和表面张力会让尿液以一粒粒水珠形式排出，因此速度慢。尿量大的流出速度快，少的流出速度慢，结果排便时间都差不多。



来源：《科学》杂志2014年3月新闻报

图2.5 表面张力影响排小便和植物的形状

第二个例子是植物种子在潮湿的地方靠表面张力维持形状。例如，枯叶洋葵的种子很长，本身维持不了直立状态。但在潮湿的空气中，表面水产生的张力使种子是直的。遇到干燥空气后，失去水分，表面张力消失，就卷起成螺旋形状。

压出水痕生浮力

我们都知道物体入水能产生浮力，其实物体压在水面上也能产生浮力。

把一块比水轻的物体放在水上，物体会浮在水面，这是因为水给物体产生的浮力能平衡物体的重量，使物体不会沉下去。其实比水重的物体放入水中也有浮力，只是浮力相比于重量不够，物体会沉下去。

阿基米德洗澡时琢磨出了这种奥秘：浮力为入水物体排开的水的重量。原来，水下有水压，是地球引力导致的，如同大气有气压。水深1米的地方的水压，正好能托起这个地方的上方1米深的水的重量。越深

的地方需要托起的水的重量越大，因此水压越大。故水压随深度成正比。

把物体放入水中，越深的部位感受的水压越大，因此会产生一个向上的力，就是浮力。由于水压并不认识你是水还是什么其他物体，因此放入水中的物体的表面任一点感受的水压，与原来占据的水在这些位置的水压是一样的。既然这些位置的水压能托起这些重量的水，那么也就能托起同等重量的其他物体。因此，把这部分水用一个物体替换，水压对这个物体的作用还是相当于原有水的重量。也就是说，浮力等于物体排开水的重量。

物体浸入水中会产生浮力是广为人知的现象。如果疏水的小物体压在水上，使水面被压出漏斗，也相当于排开了一部分水。虽然物体的受力是直接通过表面张力，但力的大小还是等于漏斗排开水的重量。这个现象就不为人熟悉了。因此，要让小物体受到的力越大，就需要压出的漏斗体积越大。稍后介绍的水黾，就是靠疏水的腿压出漏斗形压痕，产生让其能立在水面上的浮力的。

人在深水中

鸭子在水上不沉下去，是因为所受的浮力大于其重量。鸭子身上有不少毛毛，这些毛毛的比重比水的小，但压入水中却能产生不少浮力。

人类如果像鸭子一样，也就能轻易浮在水上了。可是，一部分人的身体密度恰恰比水的密度多了一点点，使我们排开水的重量即能够产生的浮力的大小，偏偏就比我们的体重小一些。

比如说，某人的身高假设是1.7米，身体近似看成直径约为23厘米

的圆柱体（当然，人的身材比这优美多了），体积近似为0.07立方米。完全沉入水中，产生的浮力为70千克，因为水的密度是1000千克 / 立方米。可是，大多数人的身体密度比水高百分之2左右。因此，这位身高1.7米、体积0.07立方米的人，体重约为71.4千克左右。就比浮力多了一点点。如果他不会游泳，就会沉入深水下。

但也有一些人可以仰躺在水面不沉下去。其他一些人可以通过脚踩水或者身体协调摆动浮在水面。通过姿势的协调、踩水或者游泳而不沉入水下，可以讲出许多科学道理，但真正用起来不是那么回事，不会那么简单。应在指导下先学会踩水和游泳，在此基础上接触科学道理或接受科学指导才可以提升技术。

据说，人直立在水中，将大腿放平，两个小腿像螺旋桨一样绕竖直方向旋转，同时让脚掌铺平，就能较容易地踩水。但也有将脚掌铺平往下蹬，往上收时竖着脚掌的踩水方法。前者可能利用了脚掌的升力，脚掌往内侧旋转运动时，看似像有迎角和弯度的机翼。后者利用下蹲过程的阻力。效果取决于协调、姿势和速度，需要有指导科学地练习。

2. 绿叶上的水珠 纯水泡泡与肥皂泡泡

植物叶暴露在有尘埃的大气中依然翠绿，上面的水珠晶莹剔透，荷叶出淤泥而不染。原来是无数微米级的乳头状小鼓包托起了小水珠，滚走任何有损容颜的尘埃。水底常常冒气泡，遗憾的是水面就是小气泡的葬身之地，我们纳闷小气泡为何不挣脱水表面而飘向空中，以便享受在空中飘扬的快乐。好像舍不得把一点点水分带到空中，一钻出水面就自毁。表面张力让水泡有了生命，可是只是短暂的生命。肥皂液似乎是长生之剂，于是我们可以欣赏五彩缤纷的肥皂泡泡在空中漂浮。

绿叶上的水滴小鸟依人 乳头状小鼓包（图2.6）

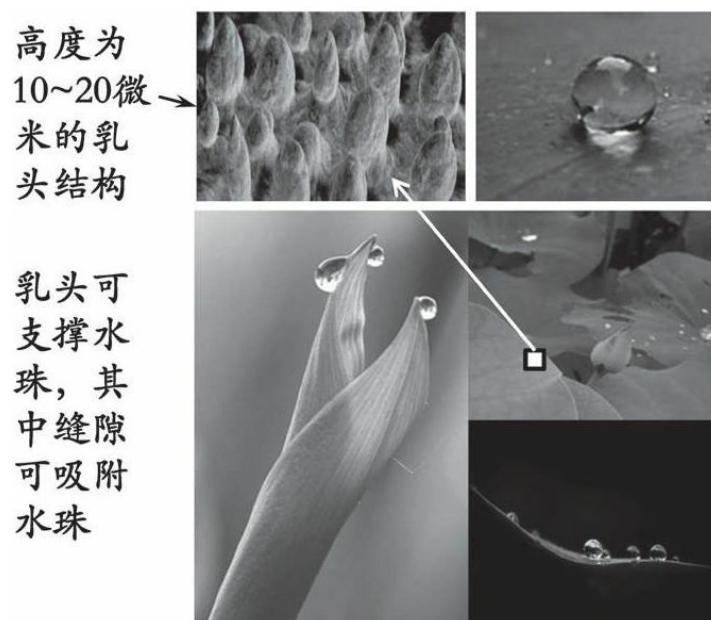


图2.6 荷叶结构与绿草上的水珠

诗人白居易在两首诗中都提到露珠。“叶含浓露如啼眼，枝袅轻风似舞腰”。“一道残阳铺水中，半江瑟瑟半江红。可怜九月初三夜，露似珍珠月似弓”。形容霜露的水珠如啼哭的泪珠，也如珍珠般晶莹剔透。水珠不仅自身美丽，还能帮绿叶美容，清除绿叶上的尘埃。

荷叶表面上长有10微米尺寸的乳头状小鼓包。水滴在上面，这些乳头状鼓包起到了支点的作用，把小水珠托起来。这如同我们用几根手指托起了一只篮球一样。鼓包之间还夹带一些空气，因此水珠与荷叶的接触面上，相当于有不少空隙。乳头状鼓包携带空气，轻轻托起水滴，于是水滴能近似维持原有的球形。表现出来，就是接触角很大。接触角大于150度（也说120度），就称为超疏水材料，荷叶就是超疏水材料。

当然，表面张力还可以把极小的小水珠粘住，使小水珠小鸟依人般地吸附在绿叶的侧边或倒挂在叶下。如果表面张力的拉扯还不够劲，那么鼓包之间的空隙就利用真空吸附即拔火罐的原理助力。小水珠留在一处概率多低，怎么会舍得水珠掉下去呢？

其实，壁虎之类的爬墙动物，脚底也是有那种凹槽波纹结构，利用拔火罐原理可以吸附在峭壁上。你可以制造某种真空吸附装置，让里面80平方厘米的范围内大气压消失，这样产生的吸附力可以高达80千克，足以把人给悬吊起来。

由于尖尖的小乳头结构将小水珠支起，小风摇晃荷叶时，小水珠就如弹球一样，溜溜滚动，顺便清洁上面的尘埃，让荷叶滴尘不染。

肥皂泡泡与纯水泡泡

用肥皂水吹出来的肥皂泡泡，五彩斑斓，一串串球状泡泡，或飘向空中，或散落在地面。碰到树枝，会轻轻砸开成小得几乎看不见、会很快蒸发掉的小液滴。肥皂泡之所以维持球面形状，也是因为表面张力，各方向紧绷肥皂泡内的空气，互不相让，最终让肥皂泡近似呈球形，以便肥皂泡表面上的形状处处一样，只有球面才可以处处一样。

我们都知道要吹肥皂泡泡，而不是吹纯水泡泡。这并不是为了显得五彩斑斓。原来，纯水的表面张力太大，吹出的泡泡很快破裂甚至吹不出来。用于吹泡泡的肥皂水表面张力只有纯水的 $1/3$ 左右。肥皂泡泡一旦吹起来，可以持续几秒钟而不破裂。另外，肥皂有一层油脂，阻止液体蒸发。如果在充满水蒸气的容器中生成肥皂泡泡，那么蒸发速度就极慢，极有可能让肥皂泡泡持续很久时间而不破裂。据说，印第安纳州的

泡泡爱好者埃菲尔·普拉斯特雷，吹出了一个泡泡，持续了341天才破裂！

有一种更科学的说法来说明肥皂泡泡应取球形，而不是椭球或者立方体形状。原来，球的面积除以体积，在所有同体积的立体图形中是最小的。表面张力也是一种能量，维持泡泡的形状也是一种代价。依据最小作用原理，应取表面积最小的球形。

3. 鸡毛掸子一样的水黾腿 镊子一样的刚毛

水面有表面张力，就像弹簧膜。弹簧膜支撑不起我们这么大的人，就像我们的弹簧卧床支撑不了大象。可是，体长只有2厘米左右，重量只有10毫克左右的水黾就不一样了。这么轻的水黾，躺在有张力膜的水面，也许像人躺在丝布吊床上荡秋千一样自由自在。

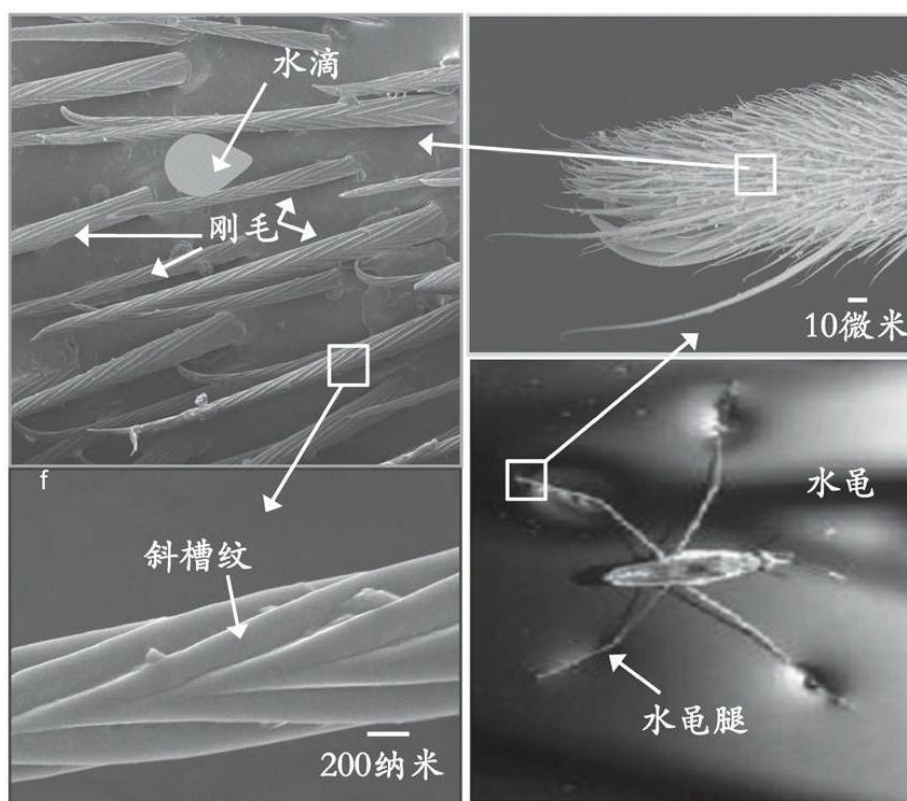
尖锥状螺纹刚毛（图2.7）



别担心水黾的腿细如针尖，它的腿不像连皮肉都可以刺破的钢针，它的腿长满了锥头状小毛毛（也称为纤毛或刚毛），就像荷叶上长满了乳头状小鼓包一样，只是小毛毛更纤细。

水黾的腿长约为2厘米，腿的直径为0.1毫米左右。中国科学院化学研究所的科学家江雷等发现，水黾腿上长满了数以千计的尖锥状的有一定弹性的小刚毛。小刚毛的直径为数微米到10余微米不等。小刚毛上有些槽纹。可以吸附一些空气，不容易让水打湿，因此带刚毛的腿有较好的超疏水性。

如同荷叶表面的乳头状小鼓包可以像手指托起气球一样把水珠托起，小刚毛可以把水面撑开。水龟的重量通过腿压在水面上，疏水的小刚毛能把有表面张力的水面往下推开，于是水面被压出漏斗型压痕。这如同我们站在弹簧垫上，弹簧垫靠弯曲下沉，来获得托起我们身体的力。有表面张力的水面，对于刚毛腿而言，下沉和产生力的方式完全与弹簧垫一样。有了许多小刚毛，就不容易刺破水面，如同我们一个大人站在弹簧床上容易压坏弹簧床，而多位体重加起来和大人体重一样的小孩站在弹簧床不同位置不会压坏它一样。



来源：江雷等用原子显微镜拍摄的水龟腿上的刚毛

图2.7 水龟腿上刻有斜槽纹的刚毛

除此之外，中国科学院化学研究所等单位的科学家江雷等在2015年报道了相邻两根刚毛的镊子功能。即像镊子一样可以夹住小雾滴，使小

雾滴无法钻进去打湿水黾腿。因此，在潮湿的地方，水黾腿也是干燥的。

压痕的深度 绰绰有余的浮力（图2.8）

2003年，中国科学院化学研究所的科学家江雷等与清华大学一组科学家合作，发现带刚毛的水黾腿横卧压迫水面时，在刚毛的憎水作用下，最深可下压7毫米左右，水面还不会被刺破。清华大学的科学家经过计算，发现被水黾腿下压的水面，向下弯曲成漏斗形压痕，压痕体积可达150立方毫米左右。压痕体积也等价于产生浮力的排水体积，即单根腿能支撑起150毫克左右的重量。实际水黾可能下压4.4毫米左右，这足以支撑10毫克左右的体重。

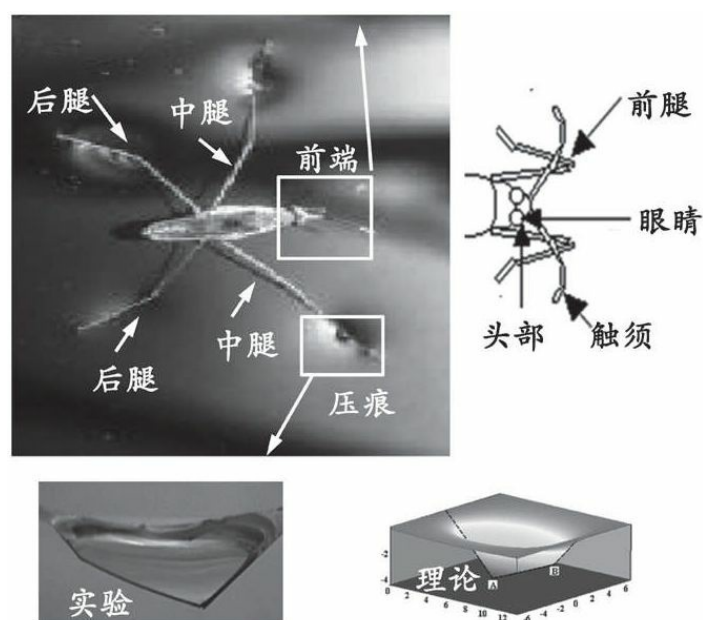


图2.8 水黾腿与压痕

于是，水黾腿可产生的最大浮力远大于其体重，多余的力可用于弹

跳或者应付其他特殊情况。例如，有风浪时，如果一个水波过来，这种能够产生的多余力可以保证水龟随波浪一起起伏，可以抵消波纹起伏时加速带来的超重现象。如果下雨，打在水龟身体正中央，那么瞬时增重后的重量比水龟自重大多了。

由于能浮在水面得益于表面张力，那么在水里放些肥皂水，让表面张力系数降低，水龟能得到的支撑力就小多了。我们可以抓住一只水龟，放在洒了足够多肥皂水的水中，看看水龟是否会沉下去。

4. 凌波微步 划桨运动 蹦床运动

甬认为小水龟是一只什么也不懂的小昆虫，它既掌握了划桨技术也掌握了蹦床技术，难怪说它能在水面凌波微步。它有六条腿，分工很奇特。两条前腿较短，用于扑食。中间一对长腿用于驱动运动。后面两条腿用于操控方向。在水面行走时最快达到每秒1.5米，比涟漪波纹的传播速度还快。它还能玩双人起跳呢。

比船划桨更奇妙的水龟腿（图2.9）

我们已经知道，橹靠迎角升力产生推进力。一般的手动划桨也可以采用迎角效应，也可以通过让桨面朝着船行进的反方向运动产生阻力来获得推进力。水龟腿如果横卧着采用划桨模式，即向后滑动来获得行走的力，则有三种可能的方式。

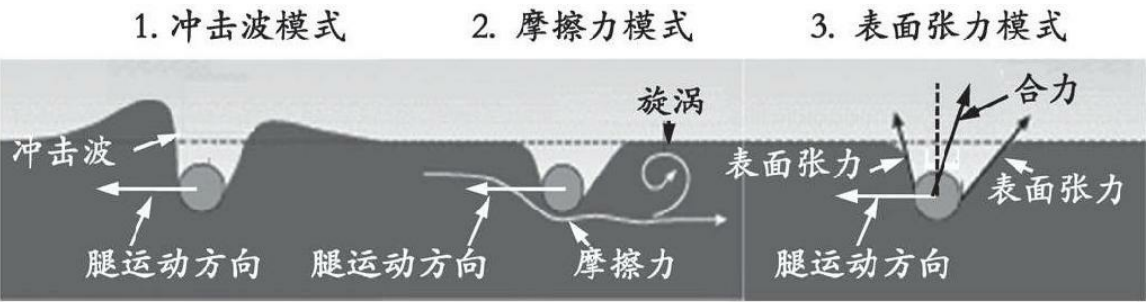
第一种是后蹬时，把后面的水位推高了，这个被推高的水位也可以称为冲击波。高水位支撑的水重大，比堵在另一侧的水位高，因此重力引起的水压差就使后蹬的那一侧水压高了，另一侧低一些，于是产生推

进力。这种模式称为冲击波模式。

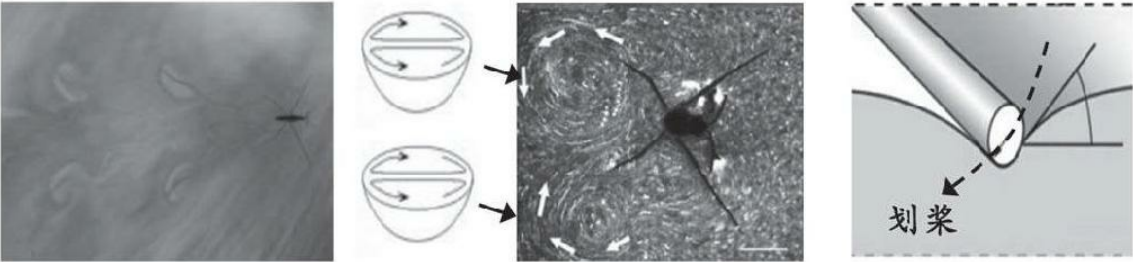
第二种是摩擦力。即腿向后滑动时，水面给腿施加的摩擦力作用在水龟前进的方向。这和我们走路获得的行走力是类似的。

第三种是表面张力模式。向后蹬腿时，后方的水面更陡，前方的更平坦。于是，与水面平行的表面张力形成的合力就有一个指向水龟运动方向的分量。

这三种模式应该都起一定的作用，合起来可推动水龟前进，可以在水面滑动行走。



《实验生物学杂志》1997年报道的三种滑动推进模式



《自然》2003年报道的划桨旋涡模式

图2.9 水龟腿滑动时推进模式以及划出的旋涡

推进力要么是压差，要么是摩擦力，要么是表面张力，要么是它们一起起作用。这是最基础的力。移动所需要的力靠中腿，后腿类似于船舵，主要用来调节方向。

人们往往用推进过程中观察到的其他现象来解释为何产生了推进力。比如说，可以笼统地这样说：借助于划桨动作，对压痕产生一向后方的作用力，使当地的水表膜向后运动，从而获得一种反作用力（前进力）推动水黾向前移动。由表面张力维持的压痕像波浪一样，因此也把这种推进模式称为表面张力波模式。用这种模式去解释水黾的推进，也存在一个困难，称为德尼佯谬，是指婴儿期的水黾划腿速度太慢，不足以以此产生表面张力波。

于是人们质疑表面张力波模式。实验观察到，水黾行走时往往会在水面留下一串串旋涡，这和船桨产生旋涡的方式有点类似。尤其人们观察到了半球形旋涡。于是人们认为是旋涡导致了推进。旋涡应该是产生了推进力的结果，而不是产生了推进的原因。旋涡的出现，表明压水较深，后方被推高的水位向后移动时，将附近的水搅出了旋涡，如同船桨划出旋涡一样。

蹦床模式 蹦床运动（图2.10）

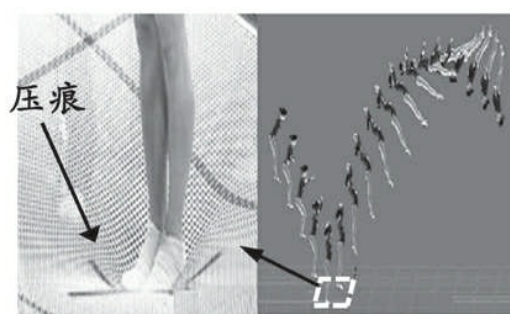
前面已经知道，单只水黾腿即可以支撑起15倍左右的水黾的重量，因此多于重量的力可以用来使水黾弹跳。

表面张力使水面像有弹性的膜。因此，小水黾可以把水面当蹦床使用。在跳动模式中，后腿可能与中腿一起在作用，除了向后蹬，还向下蹬腿。借助表面张力产生的憎水力，作用方向向上向前，使水黾能跳跃。

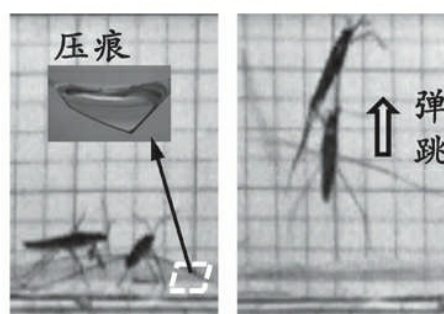
蹦床作为一个体育项目，也称为空中芭蕾运动。蹦床选手从专用弹性垫上跳起数米高，每次起跳后，要在空中完成一次规定的空翻动作，

包括前空翻、后空翻、侧空翻，也包括前空翻接侧空翻和后空翻接侧空翻。水黽在水面上也能旋转弹跳，这是由于两条中腿施加的力大小不一样造成的。

单只水黽起跳不难理解。有时能看到两只水黽一起起跳。很难想象，它们是如何沟通的，是有意识地一起起跳，还是一只起跳时把另外一只带上了，还是一只看到另外一只起跳后迅速起跳？如果是有意识地一起起跳，它们一定有一种信息沟通方式。它们肯定不会说话，但一定有通信方式。



中国蹦床网



《科学》杂志2014年

图2.10 蹦床运动与水黽蹦跳

5. 驱赶、威慑和求偶 通信与存储

小池塘、小河或者淡水湖水面上，有时会出现一圈圈的小波纹，称为涟漪，那是体长约2厘米、重量约为10毫克的水黽在凌波微步。涟漪水面波纹一高一低，水黽感受到这种水位高低的变化，能完成一定意思的信息沟通。这种信息沟通的方式，居然和现代通信及计算机信息存储具有相似之处。借此，我们正好浅显地去了解一下现代通信与存储的秘密所在。

水黽的驱赶、威慑和求偶信号（图2.11）

一只雄性水龟向其他水龟领地（水）发起攻击，它通过激发频率为25赫兹（即每秒产生25个圈圈）的涟漪来驱赶其他水龟。这个25赫兹是水龟界的驱赶信号。其他水龟如果激发10赫兹的涟漪，就表示抵抗。这个10赫兹是水龟界的威慑信号。也许，它们根据对方信号的强弱（即涟漪波纹的强度，如波纹尺寸，传播快慢），来判断对方是否强于自己，以决定是否对抗或逃离。雌性水龟如果留在原地，那么雄性水龟可能激发3赫兹的涟漪，表示友好与愿意为伴。

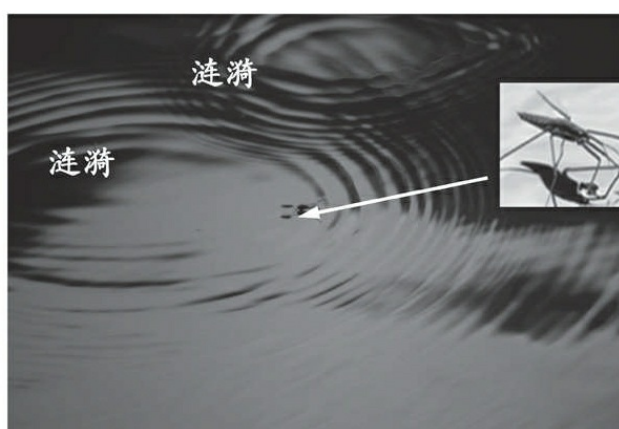


图2.11 水龟行走产生的涟漪

科学家通过观察研究，发现了水龟利用这三个频率，分别达到驱赶、威慑和求偶的作用。其他动物也许存在相似的或者完全不同的通信方式。

人类因为有了语言，因此主要通过语言（包括文字语言）方式通信。直接对话或者托人带话，是最简单最直接的通信方式。书信、电子邮件是一种比较正式的通信方式。古代靠信鸽、快马（日行千里马，八百里传书可一日送达）、靠烽火台狼烟等实现快速通信。现代通信与水龟通信之间，也存在某种相似性。

水龟激发的涟漪中的一圈圈波纹，就是水位在一高一低地变化。一

高一低的变化，只涉及了两个状态，因此这种通信方式非常简单。如果将水位变高记为1，水位变低记为0，那么理解水龟的通信就只需要与0和1打交道了。人类现代的通信以及计算机存储还真的就是与0和1打交道。

借此，我们就去看看一下现代通信和计算机存储是如何与0和1打交道的。

两楼之间的通信与存储

现代通信需要快速传递大量数据，基于的原理却较为简单。为此，先举一个形象的例子来说明，即南楼与北楼之间传递与存储信息。假设你用10110表示“停水了”，用10111表示“有暖气了”。

A楼用红绿两盏灯的切换传信号。在一个规定的时间段内（比如说5秒），第1秒内用亮红灯表示1，第2秒期间用亮绿灯表示0，第3秒期间用亮红灯表示1，第4秒期间用亮红灯表示1，第5秒期间用亮绿灯表示0。这样，就把10110这个代表“停水了”的信号发送出去了。当然可以通过时间停顿或亮黄灯，表示要开始传输信号10110，以及结束传输10110。结束后，又可以按类似方式传送10111。

B楼通过某种方式识别到A楼那边信号灯的亮灯信息，看到红灯亮时记下1，绿灯亮时记下0，通过按顺序收到红绿红红绿，于是就识别为10110，接着还可以收到10111。按照约定的字符串对应关系，就得到了“停水了”和“有暖气了”的信息。

现在看B楼如何用最简单的机械、电子或光学方式存储这些信息。比如说，要存储“停水了”，就用一系列开关来存储10110。可以这样

做：B楼针对每组数据，启用一组5个磁针来存储。让磁针指向上方表示1，指向右方表示0。收到A楼红灯信息时，用某种感应方式自动让磁针指向上方，收到绿灯信息时，指向右方。

于是，一组5个磁针按顺序分别指向上、右、上、上、右，就表示存储了10110。用另一组5个磁针，用类似方式存储10111。

B楼的人依据编码对应表，就从10110或10111可解读出“停水了”或“有暖气了”。

实际无线和有线传输以及计算机对信息的存储，依据的原理与A、B楼之间的上述通信与存储，非常类似。

近代与现代通信 编码与比特（图2.12）

电缆、光纤和无线传输是现代通信和传输的有效手段。通过编码来表示文字、数字、音频、图像和视频等。

莫尔斯电码就是最早采用的适合发送文字和数字的编码。英语单词是由字母组成，例如，but由b，u，t三个字母组成。之前发送电报时，通过长短信号的组合来表示一个字母。电报机电键按一个长信号（按键开关让电流长一点，使对方接收到一个长的声音），再连续按三个短信号，就代表了B。按两个短信号加一个长信号，就代表U。按一个长信号，就代表T。对方按顺序收到B，U，T对应的编码。接着按对照表就可得到英语单词。

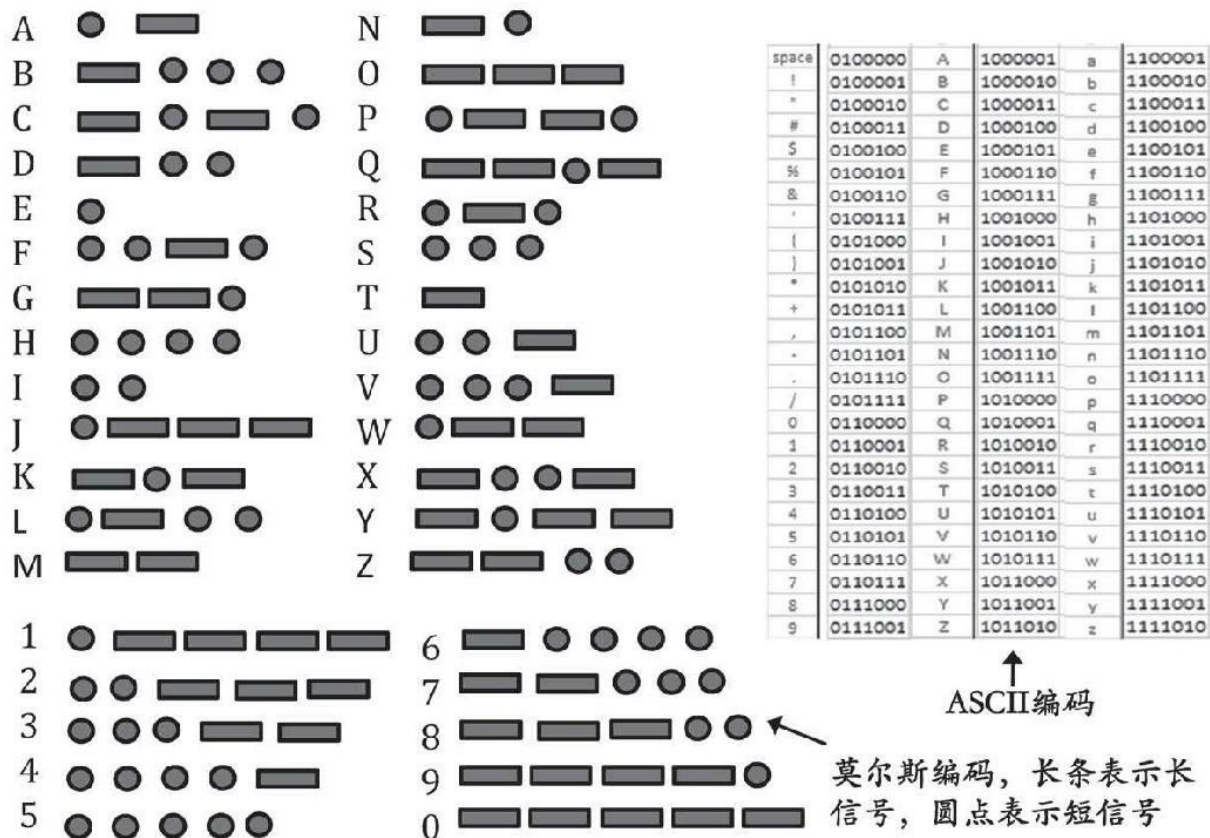


图2.12 莫尔斯电码与部分符号的ASCII编码

现代计算机则用比特来表示字母、数字、音频和视频。用1和0表示比特（bite），即无论什么符号和数字，只会涉及0和1。用一定数目的0和1排列出一个符号或一个数字。例如，用1000001这个排列表示字母A，用00000011表示数字2。

在看具体如何排列之前，首先说明一下，只有0和1时有什么方便的地方。由于只涉及0和1，因此可以用某种形式的脉冲信号（如电流脉冲）表示1或0来传输数据（例如，用正电流脉冲表示1，负的代表0）。这样，通过一组脉冲，就可以把若干个0和1组成的字符串进行传输。发送完一组，又启动另外一组（需要有规定的起止信号，以便一组字符串能被单独识别）。

一组的字符串被接收装置接收，一方面可用某种开关方式存储，另一方面用反编码转换成原始信息，例如将1000001转换为A。

在电脑中，可以用诸如电子开关的两个位置（当然还有其他方式，如两个磁极、两个不同电压、两个不同的光强、两个不同的电流）分别表示1和0，对每一个符号或数字，用一组这样的开关来存储相对应的0和1组成的编码。道理如同前面A、B楼通信用的磁针指向的存储方式。

符号、数字的比特表示

ASCII标准采用7位格式表示文字符号。例如，用1000001表示字母A。用0110001表示符号1（而不是数字1），用0110000表示符号0（而不是数字0）。我们按键盘上的A键，计算机磁盘上存储的就是1000001，当然显示时又显示的是A，即显示时，计算机又把1000001显示成符号A。

对于数字，用二进制（而不是我们熟悉的十进制，十进制会用到0，1，2，3，4，5，6，7，8，9这10个数字），即只涉及数字0和1，从而与比特直接关联。

因此，无论是符号还是数值，均可用比特表示。由于只需要传输和存储0和1，因此就非常简单。

依据不同精度要求，二进制表示的数可以采用8位，16位，32位和64位，也称为字长为8位，16位，32位和64位。以8位二进制为例，数字0表示为00000000，数字1表示为00000001，数字2表示为00000011，即相加等于2时，就进1（而十进制是相加等于10时进1）。字长为8位的计

计算机（称为8位计算机）能直接对按十进制不超过 2^8 （即256）的数进行运算，16位计算机能直接对按十进制不超过 2^{16} （即65536）的数进行运算，32位则能对不超过 2^{32} （4294967296，即42.94967296亿）的数运算，64位则能对不超过 2^{64} （ 1.8447×10^{19} ）的数运算。然而，无论是采用多少位，都有比可直接运算的数大的数。对于更大的数的运算，需要按某种方式拆分成一系列运算。

音频和视频

任何形式的声音，包括语音、音乐和背景声音，都通过空气的气压随时间的变化来传到耳朵里。记录离散时刻的气压大小对应的数，对每一组数用上述二进制的方式传输与存储。最后，通过反解码将对应的数的大小转换成压力大小，作用在能发声的喇叭上，就可以重现音频。

图像则可以分解为平面上许多点的位置数据以及颜色数据（不同颜色也可以用数值来表示），视频则还包括了时间信息。

也就是说，无论是声音、图像、还是视频，都编码成二进制数据，按前面类似方式进行传输、存储（只需要传输0和1的组合）。按某种约定的顺序或方式，进行反编码后，即可播放或显示。

可以这样简单理解一幅图像如何处理。假设你有一幅画，那么可以把这幅画所在的平面分割成数以万计的小方格（像素），每一个的中心位置信息，以及每一格的颜色信息，都用二进制数据表示。这些数据传输并保存到电脑里，调出时，就依据这些位置与颜色信息进行还原。

信息的压缩

音频和视频的数据量大，对传输与存储要求高。因此，往往需要对信息进行压缩。例如，如果一个音频只是一段正弦波，那么只需要发送该正弦波的频率、振幅和起止时间就够了，而不需要用成千上万个时刻的气压来表示。接收到正弦波信息后，就可以反算出各时刻的压力。

对于我们听到的声音，一般不是单一的谐波，而是可以看成有许许多多频率不同、振幅不一、起始时间不一的谐波叠加而成。数学上有一种变换，叫傅里叶变换，可以从一个随意得到的音频数据，提取出组成的它的各种谐波的频率、振幅等。频率太高太低的谐波，耳朵听不到或听起来不舒服。这样，就可以把那些不好的或者没用的谐波滤掉，保留那些有用的谐波。有用的谐波个数一般不太多，因此，只需要发送这些有用的谐波信息（即谐波数量，每一个有用谐波的频率和振幅，起始时间）。这样，发送和存储的数据量就小。需要播放时，解压装置就依据这些谐波的频率和振幅等，叠加生成气压随时间变化的曲线。

对于图像与视频，也有相似的压缩技术。与字符采用ASCII编码不同，图像采用JPG（JPEG）等格式编码，音频采用MP3（MPEG-3）等格式，视频采用MP4（Mpeg-4）等格式。MP3和MP4包含了某种方式的压缩技术。

2.2 风蚀地貌 圆与旋涡的生命力


弯弯的河道、凹凸有致的山形、浪涛一样的沙丘、羽扇般的翅膀，是自然演化过来的。自然演化的物体或现象往往具有奇特或优美的形状，这是因为演化遵循一些最小作用原理，优美的形状在风蚀水洗时相互作用最小。棒棒糖在舔食时的消融，也是一个类似的过程。受水流冲刷的一侧往往具有优美的弧线外表。这是与水流流线形态自适应的结果。孩提时代就会问：棒棒糖舔多少下就能舔到中心？局部拥有圆弧一样的表面形状，往往是演化的最终结果。与圆相关的圆周率看来不单单是一个神奇的数学常数，也是一个神奇的物理常数。比如说，旋转的澡盆涡是绕圆周旋转，而不是绕不规则的路径旋转。

1. 棒棒糖舔一千下 地貌的演化

儿童也有一个长期悬而未决的问题：棒棒糖需要舔多少下，才能舔到棒棒糖的中心位置？这个问题，2015年有了大致答案：舔1000下左右，即可接近棒棒糖的中心！虽然如此，得出这一答案的科学家还是认为1000下很难验证。

但这个问题牵扯出了更一般的有趣问题。例如，在水流冲刷下，堤岸、地貌、山丘等各具形状，形成了所谓的风景与地貌。糖块、药片和化学材料也会加速消融。这个过程包含了侵蚀、磨蚀、冲蚀、溶解和融化等化学物理过程。这些过程与我们的日常生活以及我们看到的大好河山息息相关。这里面应该与舔食棒棒糖一样，有一致的诀窍。

几何形状（图2.13）



我们在童年舔食棒棒糖。棒棒糖有的是球形，有的是圆盘形，有的是尖锥形，有的是圆柱形，有的是立方体，当然可能还有其他各种怪异形状。由于经常涉及一些几何图形，因此这里归纳一下常见的形状。

平面图形中有三角形、四边形和多边形。四边形中还有正方形、梯形和菱形。正方形是规则平面图形的一种，其他规则多边形有正三角形、正五边形、正六边形等。圆可以内接正多边形，当边数趋于无穷大，正多边形就是一个圆了。非规则的多边形就数不胜数了。

总之，有无数种多边形，包括有无数个正多边形。这些多边形可以作为立体图形的面，这样的立体图形称为多面体。

对于球与椭球这样的带曲面的立体图形，可以想象成也是由数目无穷大的多边形作为面围成的，就像用边数趋于无穷大的内接正多边形可以逼近一个圆一样。对于六面体，如果每一个面都是正四边形，那么这个六面体就是正六面体。正多面体的每一个面都是正多边形，所有的棱边一样长。

不知你会不会受圆可以用正多边形逼近这样一个事实的启发，想当然地用内接正多面体逼近一个球，比如说正一万面体。然而，非常奇特的是，没有正一万面体。正多面体只有五种：正四面体，正六面体，正八面体，正十二面体，正二十面体。这五种正多面体，每个面要么都是正三角形（正四面体，正八面体，正二十面体），要么是正正方形（正六面体），要么是正六边形（正十二面体）。这是因为，只有三角形、正方形和正六边形，才能进行无缝对接。

非规则的立体图形数目之多就无法描述了。单是雪花的形状，就不胜枚举。自然界很难找到完全一模一样的两片雪花。

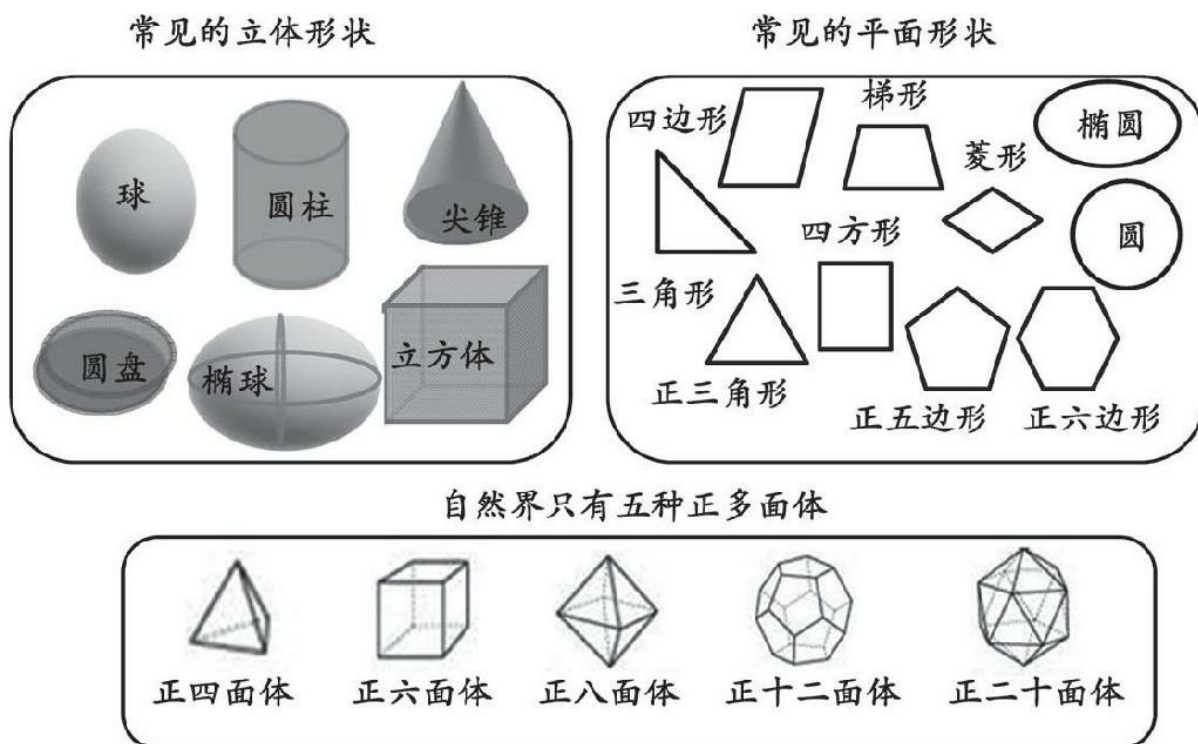


图2.13 常见几何形状：平面图形与立体图形

圆球和椭球虽然是弯曲的表面，但属于规则形状。液滴越小，越接近球形，因为表面张力试图让面上各点都一样。不管形状如何，如果水流和气流通过摩擦和溶解能改变物体的外形，那么什么形状容易顺应水流或者气流，什么形状就可以维持到最后。

棒棒糖的溶解（图2.14）

纽约大学库朗特数学研究所的博士生黄（Jinzi Huang）以及佛罗里达州立大学数学系的莫尔（Nicholas Moore）教授在纽约大学应用数学实验室将糖果放在水中任由水流冲刷。结果出乎意料：无论糖果初始形状如何，或球形或圆柱形，无论水流速度如何，最终会被冲刷溶解成一个特有的一致形状，这个形状出现后，将维持该形状，直到最终消失。

这个特定的形状在水的上游一侧近似为球面，下游一侧则被旋涡削平。原来，这种旋涡流动易生成湍流，湍流通过脉动轻易地把高浓度的溶解糖分带走，把清水送过来继续溶解。要知道，糖更容易在含糖浓度低的水中溶解。

好像水特别喜欢吃糖，只有通过削平下游表面、激发拐弯涡和湍流，才能让附近的清水跳进旋涡舔食和溶解糖分。

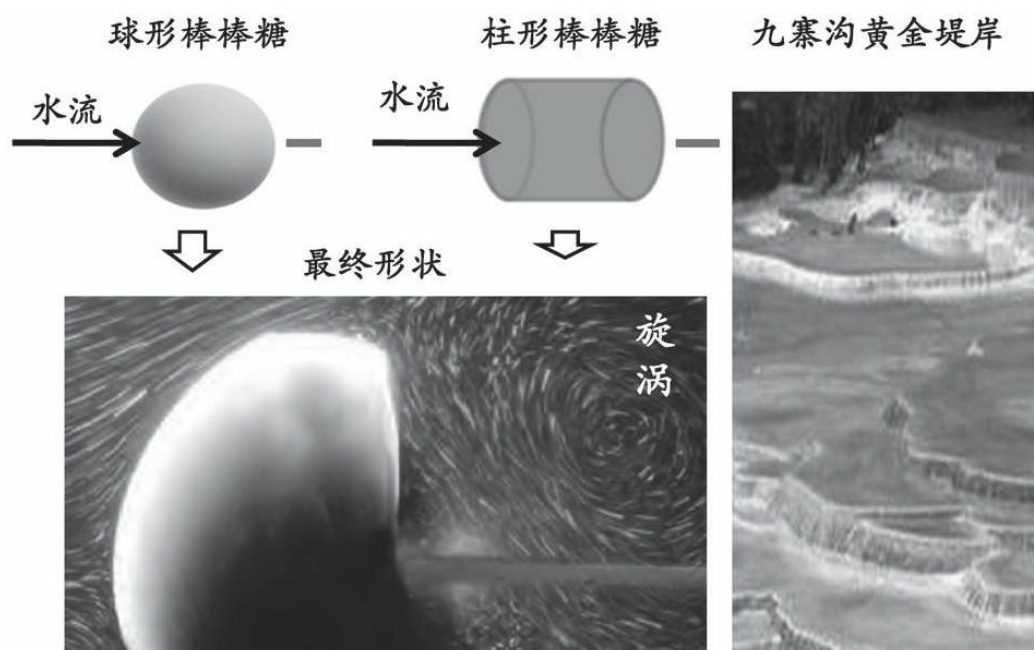


图2.14 棒棒糖的最终形状与黄金堤岸的形状

因此，直接被水流和气流冲刷的面，容易被水流打磨成光顺带弧度的球面一样的形状。这样，朝前看形状优美，面子上好看。在下游一面，却在用最有效的方式抢食棒棒糖。

最终一致形状的出现，也许能解释山形地貌如何形成。长年累月的水流冲刷，会让岩石和堤岸接受水流冲刷的一侧变得光顺且带点弧度。九寨沟翠海迷人，如梯田般的形状，黄金堤岸在水浅时露在外面，水深

时埋入水中。日积月累的水流冲刷，使堤岸如同打磨过，立体般的画卷，是水流起到了雕刻师的作用。

2. 河道弯曲 牛轭湖 蜿蜒度

黄河干流多弯曲，俗称九曲黄河。浏阳河，弯过了九道弯，也是这个意思。自然形成的河道，总是弯弯曲曲，哪怕是在平原地带也是如此。长长的河道的两岸，长年累月经奔腾的河水冲刷，理应变得更直，这样水流才会顺畅。可是，河水非得绕那么大的弯子。弯到一定程度，甩出去形成圆环型牛轭湖。这种舍直取弯的演化，给岸边建筑和跨河修桥也带来麻烦。当然，也不是毫无规律的弯曲，形状不仅像抖动的弦，弯曲程度还能跟美丽的数学常数拉上关系。

旋涡的雕刻作用（图2.15）

一条直的河道不可能是被推弯的吧，这么大的蛮力从何而来。弯曲的形成有不少因素共同起作用，其中就包括这样一个可能的机制：一旦稍有弯曲，会导致水流走弧线，引起的离心力指向弯曲的一侧，从而驱动横向流动，将两岸的泥沙交换位置，且朝着加强弯曲的方向。

假如河道首先是直的，那么由于水流不均匀或者河岸土质流失，偶尔会导致局部河道产生弯曲。如果出现这种情况，河水流动时，顺着弯道就产生由河道凹的一侧（内岸）指向河道凸的一侧（外岸）的离心力。水面的水沿河道的流速比河底的更快（河底的摩擦阻碍了流速），因此离心力更大。这种离心力使河面的水由内岸横着流向外岸。

由于河床摩擦对水流的阻碍作用，因此河床水流速度低，不会产生

很强的离心力，于是在河底不存在由内岸指向外岸的流动，相反，河面的横向流动在河床产生回流（水流必须循环起来，否则不知道去哪里了）。于是，在外岸，这种回旋流动就把河岸的土质向河床方向冲刷，使外岸泥土进一步流失，向外拓展。冲走的泥土被回流输送到内岸，使内岸进一步向外岸方向拓展。这也许就是越来越弯曲的一种原因。

总之，一旦某处有弯曲，就会在弯曲的地方产生离心力，形成横向回流，回流的冲刷使河道进一步变弯。当然，可能还有一些其他原因。有的用地球自转线速度在不同纬度的差异带来的离心力差异和水位高差异来解释。河道弯曲的演化过于漫长，很难检验哪种原因更正确。

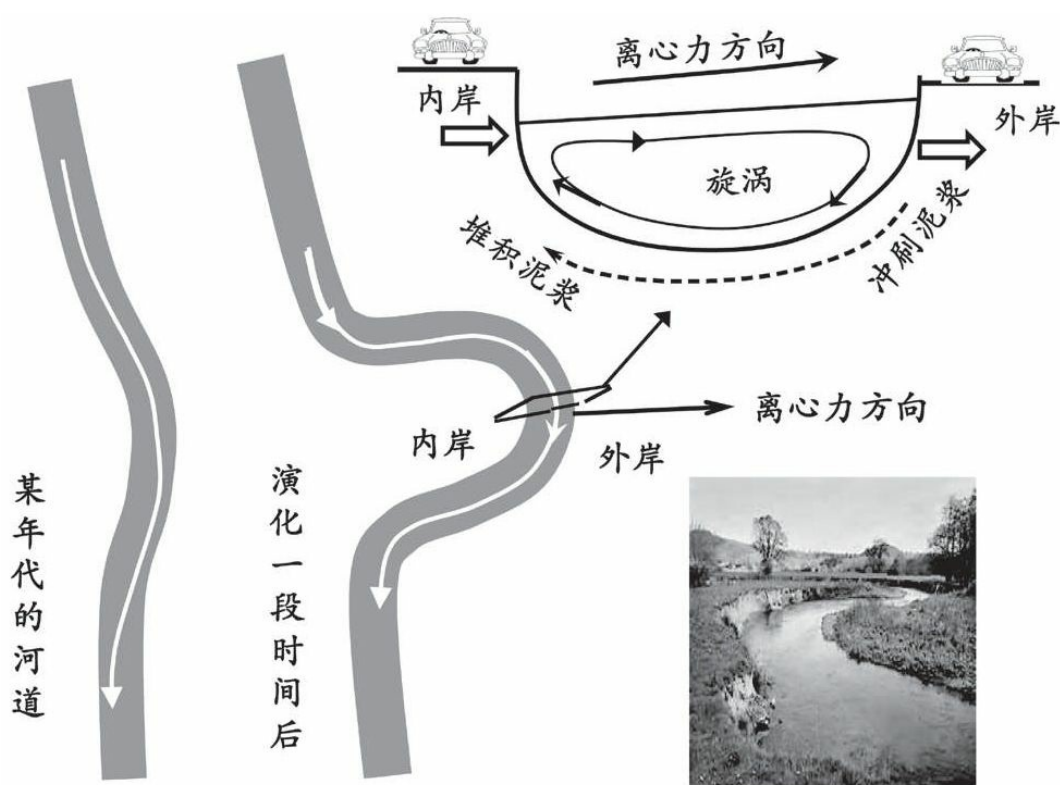


图2.15 河道走弯的一种原理

截弯取直生成牛轭湖（图2.16）

似乎没有理由阻止河道越来越弯，最弯莫过于形成圆周形河道。

河道如果出现大的拐弯，就有可能接近绕了一个圆周。这种局部圆环形河道的入口和出口如此接近，以致在洪水或其他因素作用下，可能直接打通。

如果发生这种情况，那么一方面使河道截弯取直，另一方面将原有的局部圆环形河道（近似牛轭形状）阻断，形成牛轭湖。河道边上常见的圆环形小湖，就是这么来的。



图2.16 局部弧形河道截弯取直形成牛轭湖

蜿蜒度与圆周率（图2.17）

河道弯曲程度可以用河道长度与两端直线距离的比值来衡量。该比值称为蜿蜒度（Sinuosity）。黄河是中华文明的母亲河，干流的河源在青海省的玛曲（青藏高原的巴颜喀拉山脉，北麓约古宗列盆地），经青

海、四川、甘肃、宁夏、内蒙古、陕西、山西、河南及山东流入渤海。全长为5464千米左右，是河源与出海口之间的沿地球表面的直线距离的2.64倍，因此蜿蜒度是2.64。长江干流发源于青海格拉丹东，止于上海长江口，沿地球表面的直线距离约为2875千米，蜿蜒长度约为6397千米，故蜿蜒度约等于2.23。

1996年，剑桥大学汉斯-亨里克·斯托伦（Hans-Henrik Stolum）教授在《科学》杂志上发表了一篇论文，将河道弯曲看成某种演化过程（一种自组织过程），用数学上的所谓分形几何理论，证明蜿蜒度的理论值应介于2.7和3.5之间，平均值正好等于圆周率 π ，即约为3.14。

于是，人们产生了这样的常识，即一条蜿蜒流淌的河流从源头到河口之间曲曲折折的总长度平均是其源头到河口之间（沿地球表面的）直线距离的 π 倍。后来，杰姆斯·克里姆比较了258条河流，发现蜿蜒度平均值比这小，只有1.94。但对于平原地带的河流，蜿蜒度据说还是比较接近圆周率。黄河蜿蜒度是2.64，小于3.14但大于1.94。

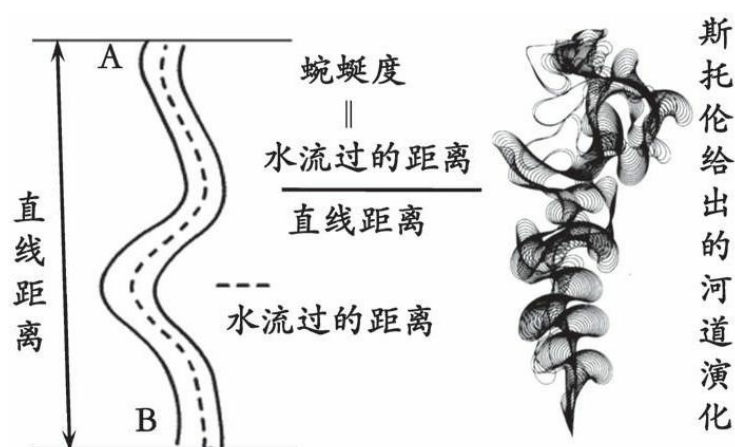


图2.17 蜿蜒度与不同时期的河道形状叠加图

3. 圆周率 自然界中的常数 数学的魅力

弯弯的河道，蜿蜒度的理论值居然和圆周率联系在一起。圆周率本身是圆的周长与直径的比值。圆周率是一个常数，无论圆是多大，这个比值都是这个常数。圆周率的常数性让人们想到：复杂的自然界，是否由类似于圆周率这样的简单数学规律构成。我们首先会问，这样的常数是独一无二的吗？于是，欣赏圆周率之前，我们有必要看看一些别的常数。

奇特的常数（图2.18）

真空中的光速是个常数，不管是什么发射的光，不管是蓝光红光还是紫光，在真空光速都是299792458米 / 秒，简单地记忆就是每秒30万千米。令人不可思议的是，不管发出光线的物体以何速度运动且不管测量者以什么速度运动，测得的真空光速都是每秒30万千米。另外，引力波、电磁波的速度也等于这个光速。

温度再低，也不可能低于-273.15摄氏度。这是绝对零度。给任何物质降温，都不可能低于这个温度。

向上，不能超过光速，向下，不能低于绝对零度。人们试图用各种方式去解释为何不能超过光速，但往往把从光速不能超过这个假设附带得到的结论，再用于去解释光速为何不能超过。

普朗克常数（约为 $6.62606957 \times 10^{-34}$ ，单位为焦耳×秒）也是值得一提的量。光线是由光子构成的，一个光子的能量等于其频率乘以普朗克常数。

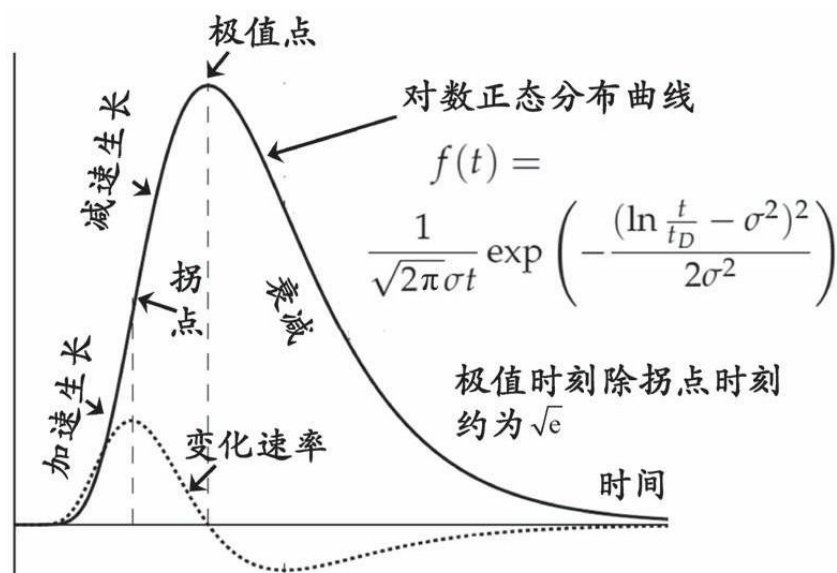


图2.18 一类生长衰减问题的演化与 e 相关

这几个常数有一个特点，如果改变单位制，那么这些常数的取值就不一样。比如说，真空的光速，如果以每小时多少千米作为单位，那么就是每小时10.8亿千米左右。

还有一个奇特的常数，称为宇宙精细结构常数，用符号 α 表示，大小等于137.03599913，没有单位。理论物理学家、宇宙学家和天文学家保罗·戴维斯在2016年2月专门发文讨论这个常数。它是普朗克常数乘以光速，再除以电子电量的平方，最后除以两倍的圆周率，得到的数。这个数在物理学中有重要地位。比如说，电子激发后发射光子的速度快慢就是由这个数来衡量的。由于这个数没有单位，因此它就是那么大。科学家一直搞不清这个常数是怎么来的，甚至怀疑宇宙中有更深刻的理论。数学家也试图找到一个公式，让这个常数与其他更好理解的数（如圆周率）关联起来。物理学中的认知规律（比如说牛顿定律）可以是常识与直觉，但数学基于逻辑推演应能给出严谨的结论。可是，这些努力都没有搞清这个常数的来源。

据保罗·戴维斯说，韦布用某种方式测量了十亿光年以外天体发出的谱线对应的宇宙精细结构常数，居然得出了这个数有可能不是常数的结论。如果此结论得到确认，那么整个理论物理学的根基将受到威胁。

有一类先生长后衰减的现象，例如流行病的增长与人口增长率。随着时间的推移，先加速增长，接着出现拐点，即增长率放缓，但还继续增长。到了某个时刻，不再增长，出现极大值，随后衰减。这种过程的变化规律往往是对数正态曲线。

据期刊《熵》（*Entropy*）报道，对于大多数问题，极大值发生的时间与拐点发生的时间之比值，理论上应等于根号 e （ e 约为2.718），即约等于1.6487。这个 e 也是一个神奇的常数。艾力·茂尔（Eli Maor）2009年就写了一部由普林斯顿大学出版社发表的书，书名为“ e ：一个数的故事”。可见这个 e 有多重要。学习过一点微积分的都知道，用这个 e 可以构造一个指数函数，其微分总是等于这个函数自己，不管微分多少次。

圆周率与科学家日

现在可以回到圆周率了。圆周率（符号为 π ，读为Pi）可以说既是一个常数，也是一个与单位无关的常数，尤其是可理解的常数。也许没有一个物理学和数学的其他常数可以给科学家的节日命名，而每年的3月14日被定为圆周率日，也称为科学家日。一般的庆祝活动包括吃苹果派之类的派（因与 π 谐音）。

圆周率前三个数字是3.14，科学家日就是这么来的。其精确的值在小数点后有无数位。如果小数点后取15位，它是3.141592653589793。

圆是一个简单优美的外形。按理，其周长与直径之比对应的数是一

个有理数（有理数就是能表示为两个整数之比的数，如0.4可以表示为2：5）。可是，糟糕的是， π 却是无理数，也就是说它不能表示成两个整数之比。音乐往往由乐器发出的具有不同振动频率的声音叠加而来，如果不同的频率具有整数比，会比不是整数比的动听一些，乐感和谐一些（因此叠加出来的音称为和音）。既然圆这样的最优美最简单的外形是一种和谐的图形，那么为何反映周长与直径之比的 π 却是一个无理数？

不仅如此，它还是十进制下的无限不循环小数（小数点后有无数位，且不会出现循环，用1除以7，得到的0.142857142857.....就是一个无限循环小数，小数点后142857不断循环）。连毕达哥拉斯这样的数学权威都否定这种无理数的存在，声称它们与一个精心设计的宇宙不相容。

古人用圆的内切多边形来逼近圆周，以此来计算圆周率的具体值是多少。多边形的周长好计算。古代数学家祖冲之算得的圆周率在3.1415926和3.1415927之间。阿基米德用了96边形来逼近圆，得出 π 的数值在223/71到22/7之间。印度数学家马德哈瓦用他的无穷级数方法，将 π 计算到小数点后十位。现在人们用计算机能算出小数点后数万亿位。

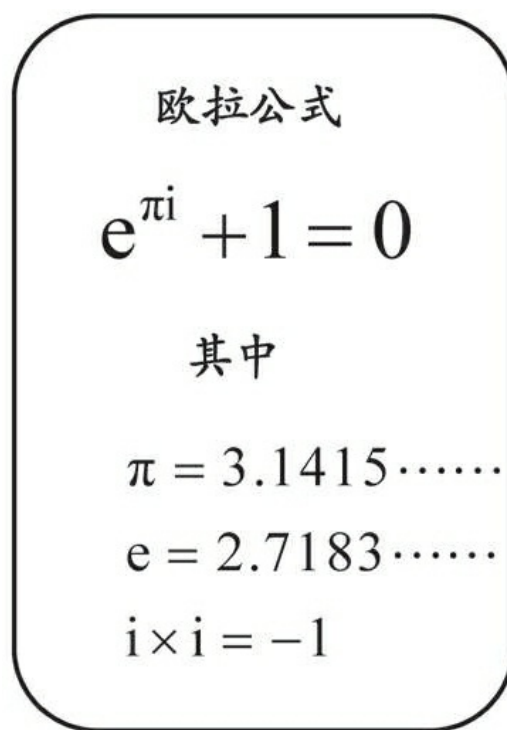
美国人口在2012年8月14日下午2时29分正好升到了3.14159265亿。为此他们展开了庆祝。用人口学家霍华德·霍根的话说：这是很多代人才有一次的的事件。

自然与数学中的圆周率（图2.19）

圆周率不仅对圆有意义，在物理学规律和数学中随处可见。例如，

机翼匀速飞行时，升力系数的理论最大值是两倍的圆周率。当然，这只是理想的最大值，实际比这小很多，虽然蝙蝠的升力系数可以接近5。

像圆周率这样的神奇的数，往往会出现在关联数字与规律的公式中。以牛顿第二定律即力（用符号 F 标记）等于质量（用 m 标记）乘以加速度（用 a 标记）为例，公式表示为 $F=m \times a$ 。用字母符号与运算符号来表示关联数字与规律的等式，统称为数学公式。数学公式简洁漂亮，可以表示各种复杂程度的数学和物理规律。



欧拉公式

$$e^{\pi i} + 1 = 0$$

其中

$$\pi = 3.1415 \dots\dots$$
$$e = 2.7183 \dots\dots$$
$$i \times i = -1$$

图2.19 最美丽的数学公式

被学术杂志《*The Mathematical Intelligencer*》评为史上最美的公式为欧拉恒等式，也称欧拉公式。它把最神奇的三个常数（圆周率 π 、自然底数 e 、虚数单位 i ）和二进制所涉及的0和1这两个最基础的数，用加法、乘法和乘方这三个最基础的运算联系在一起。

自然底数（也是无理数，约为2.718）与圆周率一样神奇，它的意义只有学习了足够的数学才能理解。虚数单位是-1开根号得到的数。我们知道1的平方就是1。虚数单位的平方就是-1。一个数（如2）的平方就是与它自己相乘得到的乘积（如2的平方是4），这个乘积（如4）的根号就是原来那个数（如4的根号就是2）。负数表示欠的数。0减1，没有可减的，即还缺1，因此把0减1得到的数记为负1，用符号表示为一

1. 虚数单位是这样定义的，它的平方等于-1。

数学的价值举例：指数与对数的放大镜效应与望远镜效应（图2.20、图2.21）

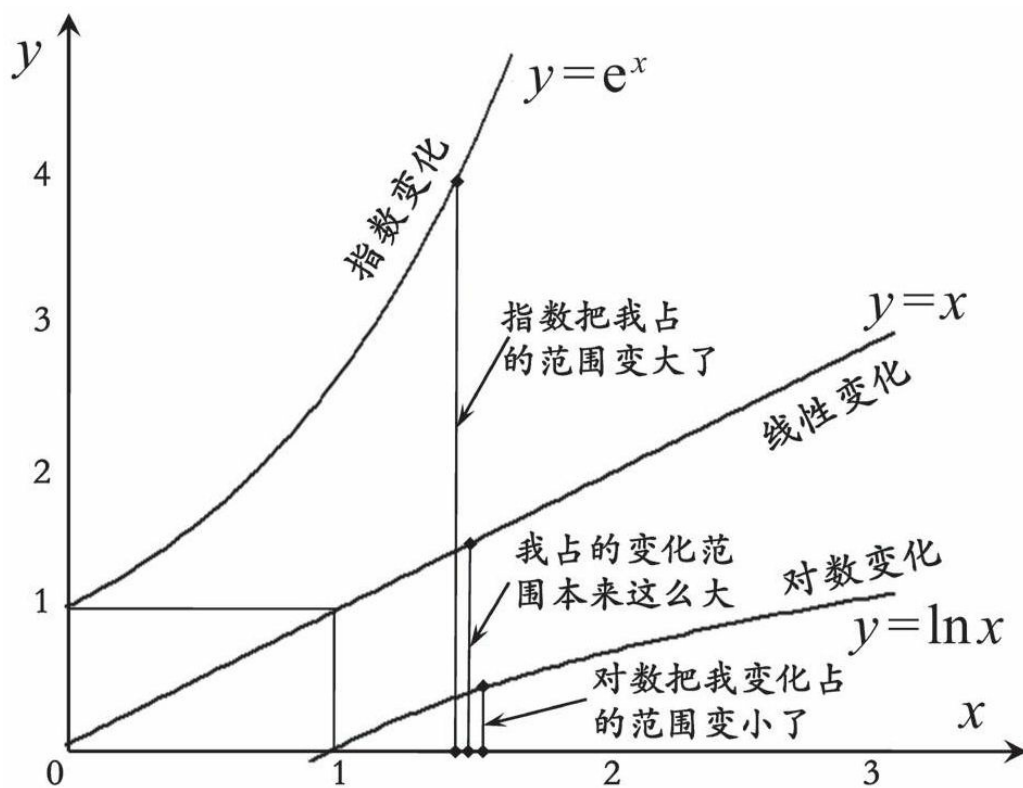


图2.20 指数函数拉宽范围，对数函数缩小范围

数学的价值不可能在这里用整个数学来说明。我们仅仅举指数与对数的例子。原因在于，我们许多地方都要用曲线来表示一些变化规律。如果曲线涉及的数的变化范围太大或者太小，那么用指数或对数，就可以把表观展示范围按要求缩小或者放大。指数和对数可以起到放大镜或望远镜作用，把大的东西缩小（望远镜）而小的东西放大（显微镜或放大镜）。例如，对于大于1的数，对数起到望远镜作用，把大的看小了或者说把远的看近了；对于小于1的数，则起到放大镜作用（虽然被放

大的数要加一个负号)。

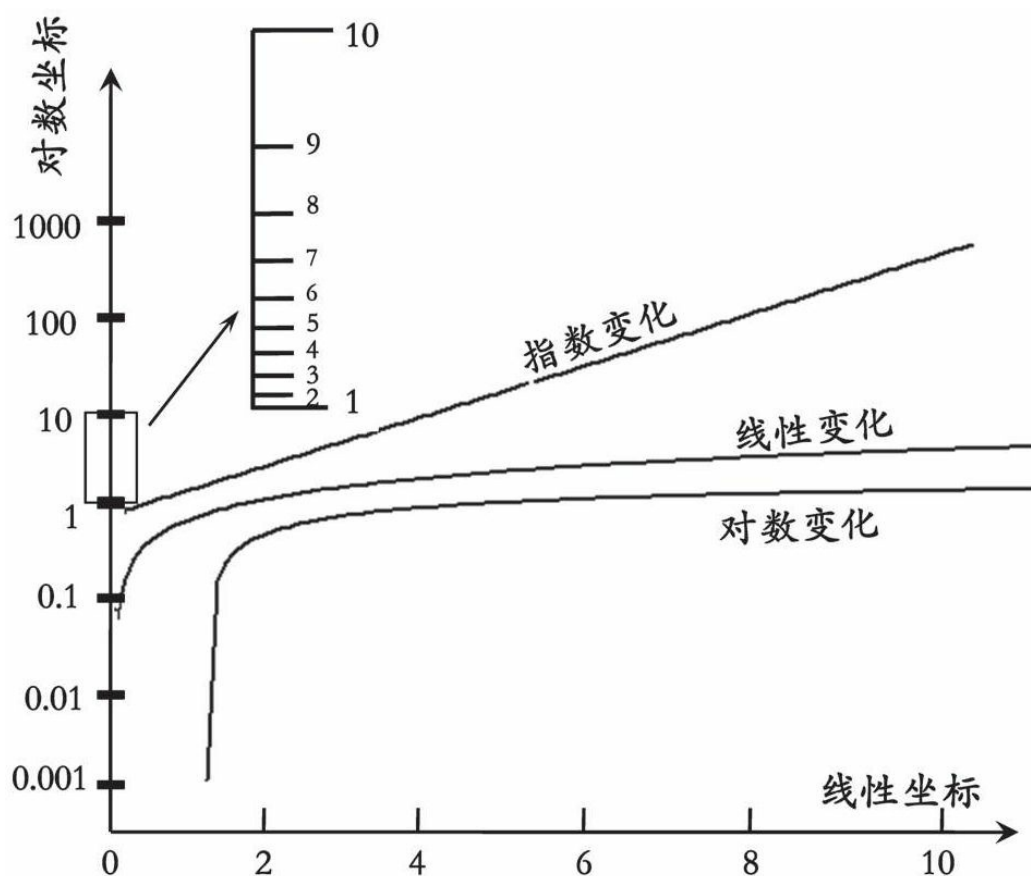


图2.21 对数坐标距离每增加一倍，代表的数增加10倍

我们常常要画曲线，要表示或展示一些结果与影响因素的关系。有的数变化范围很大，有的变化范围很小。比如说，原子数目以万亿计算。噪声所涉及的气压变化，可能是0.00002帕（牛顿 / 平方米）的量级。这记忆起来很麻烦。对数和指数表示可以让这些数被展示为一下子就能念完的数。

以10为基准为例，一个数，每增加1，其指数就放大10倍，每增加2，指数就放大100倍，如此类推。比如说，一个数增加到10倍，其指数就增加到 10^{10} 倍。

又考虑以10为基准的对数。1的对数是0，10的对数是1，100的对数是2，100万的对数是6，如此类推。即1后面有几个0，那么对数就是这些0的个数。用了对数后，原来的数的变化范围就缩小了。哪怕是10亿（9个0），其对数也就是9。

如果数小于1，那么指数就把范围缩小，对数就把范围扩大。

还是以10为基准，1的对数是0，0.1的对数是一1，0.01的对数是一2，0.001的对数是一3。如此类推，0.0000001的对数就是一7。很小的数的对数，看上去都是一个不大的数乘上一个负号。

有时为了在一张图上展示数变化范围很大的曲线，那么可以用所谓的对数坐标。对于大于1的数，纸上的坐标距离每放大1倍，那么代表的数就放大10倍。这样，纸上坐标每等距地上拉1格，那么代表的数就可以分别是1，10，100，1000，...。

对于小于1的数，纸上的坐标每等距地下拉1格，代表的数就可以分别是1，0.1，0.01，0.001，...。

如果以自然底数e作为基准，也可以类似地定义指数和对数。

数学的一般价值（图2.22）

圆周率这样的常数，在自然规律和数学中频繁地出现。这是自然规律满足数学法则的一种体现。自然规律可以总结为数学表达式，如牛顿定律。这些数学表达式可以求解，可以给出物质世界的演化规律和结果。这就是数学价值。

数学的魅力在于，一些有重大价值的物理学规律，往往可以用精美的数学表达式表示。自然规律的细枝末节可以由这些数学表达式求解获得。

我们需要用数字来表示计量、顺序、大小、位置、程度、快慢和变化等。对数的思考以及运用，诞生了推动古代与现代社会进步尤其是科技进步的数学。日常生活需要数的运用。社会发展需要数学。数与数学只有推动社会进步，才有价值。数学只有使我们达到同样目标时付出的代价更小，才有意义。上学时如果我们被数学搞怕了，这是因为数学本身与个人的目标可能不一致。虽然如此，理解一点点数学，感受其简易以及有价值的一面，可能会对我们理解世界、感受世界有所帮助。

在一般人看来，数字与数学高深而枯燥，而在数学家看来数字与数学神秘如艺术。事实上二者之间不断妥协，才使数学家发展的数学让普通人也能使用（其中的一部分）。2015年，10岁女孩晶晶因厌倦数学与两位同学一起写了“数学是死亡之源”那首诗。这说明，晶晶接触了数学枯燥与深奥的一面，让她对数字与数学失去了兴趣，甚至认为数学是万恶之源。

单是数字本身，就被伟大的数学家毕达哥拉斯认为是万物之源，更具艺术色彩的数学当然不是万恶之源。

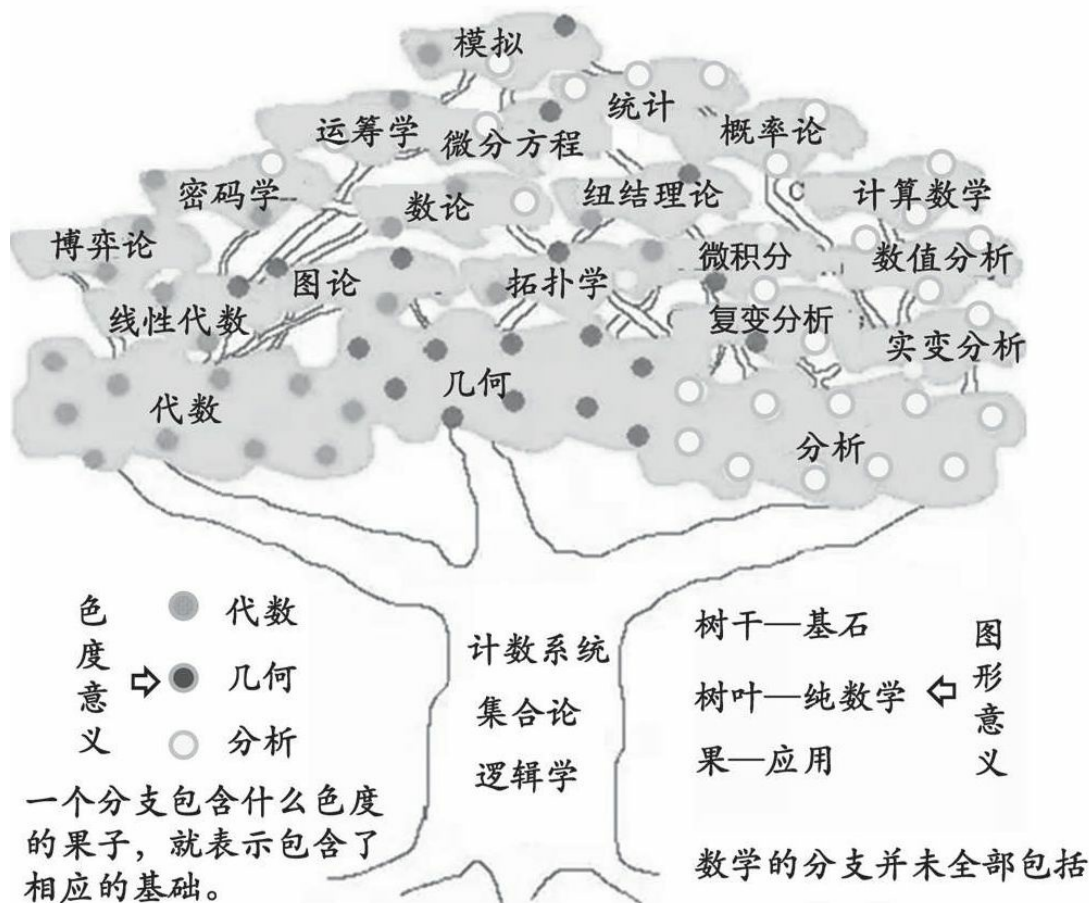


图2.22 哈勒（Hale M，史丹森大学）的数学树

数学家当然有理由把数学看成艺术，如同音乐家通过创造美妙的音乐把音乐当成艺术一样。我们普通人只有一少部分可能成为数学家，更多的是数学的使用者和数学发展的受惠者。除此之外，普通人可以通过更巧妙地利用都能掌握的数学来面对工作与生活。正是数学的发展将与数相关的事物简单化与规律化。也许正是因为数的研究与数学的发展，才导致一代代活得更久，而不是晶晶所说的数学让一代代人死得超快！数字与数学可以帮助更好地总结自然规律，也可以让日常生活更方便。

科学问题的规律用数学语言表述，构建科学规律满足的数学公式，通过求解数学公式，又能揭示更多的规律，并可为工程与社会应用提供

数据、图表和曲线。

数学以计数系统（如十进制）、逻辑学和集合论作为构建数学的基石，以几何、代数和微分作为三大基础。在中学我们就学习了不少几何和代数的知识，在大学我们会学习一些微分分析。基于三大基础，还有许多数学分支。工程技术需要用到很多数学，那是从各分支的数学中剥离出来的，可以让数学家以外的人可直接使用的数学。我们普通人在使用数学成果时，甚至都不会知道来自于什么数学分支。可见，数学家会去解决数学中的数学问题，而数学家以外的人主要使用数学家的结果。

4. 澡盆涡

澡盆涡是一个常见的日常现象，因此不但物理学家对澡盆涡感兴趣，普通人对它也有兴趣。澡盆涡是这样一种东西：它很常见，有了一点物理学知识似乎就可以大胆解释它，很容易出现在网络和科普宣讲之中；但细细思考又不是那么回事。于是，哪怕是顶级专家在写教科书时，对澡盆涡避而不谈。与陀螺仪成百上千的文献相比，澡盆涡的学术论文在正规刊物中屈指可数。难道，这么令人好奇的常见现象，在严谨的科学家心中似乎存在一些无法理解的秘密，由于很难透彻地理解而不敢轻易下结论？

澡盆涡现象（图2.23）

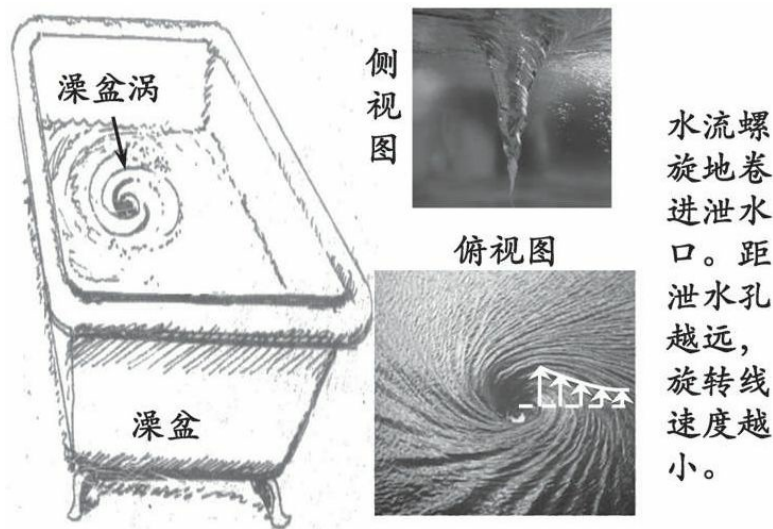


图2.23 澡盆涡常见于澡盆泄水过程

洗脸池、马桶和浴缸等放水时，往往在出水管道上方出现一个快速旋转的旋涡。这种旋涡就是澡盆涡，也称为马桶涡。任何容器中，如果底下有小孔放水，就可能出现澡盆涡。

洗脸池放水时，不是所有情况下都会产生澡盆涡。如果你试图用手掌以泄水口为中心、在水面画一圈，那么很有可能就会立即出现澡盆涡。旋转的澡盆涡的涡心是低压区，容易将空气带进去，你会听到吼吼的声响，这是空气柱与旋转的水发生摩擦产生的噪声。

陀螺之类的刚体在自转时，刚体上一点旋转的线速度与该点离开转轴的垂直距离成正比，即离转轴越远的旋转线速度越大。因此，刚体的自旋是各部分被带着一起整体旋转。

水这样的流体则很难像刚体那样被带着整体旋转。离旋转轴越远的流体，越难被带动。结果是，在以泄水孔为中心的圆周上，旋转线速度与该圆周周长的乘积，与该圆周的半径没有关系。原来，旋转也是一种惯性，上面的乘积就是旋转惯性的一种度量，没有理由改变这个惯性的

大小。由于半径越小周长越小，因此越靠近澡盆涡的中心，旋转速度越快。

除此之外，由于旋转产生的离心力需要通过往涡心方向降压来平衡，因此越往内圈水压越低，从而水位越低。除了随澡盆涡一起旋转，水还有流向涡心的运动，这种径向运动与旋转运动叠加起来，就是螺旋线运动。因此，一滴水最终被泄水孔吞下之前，走了一段螺旋线。这与螺旋电钻打洞的过程有点类似，也与开葡萄酒的螺旋起子的运动也类似。

地球自转导致的南北半球的澡盆涡（图2.24）

有人做过实验，发现在地球北半球自然产生的澡盆涡，从上往下看是逆时针的，在地球南半球，自然产生的澡盆涡是顺时针的。于是，人们认为，一般情况下也是地球自转导致了澡盆涡。

你站在北极的上面往下看，地球的自转是逆时针方向的。站在南极的上面往下看，地球的自转是顺时针的。在北半球其他地方，地球自转轴并不像北极那样与地球表面相互垂直，但这种逆时针效应还是存在的。南半球也类似。

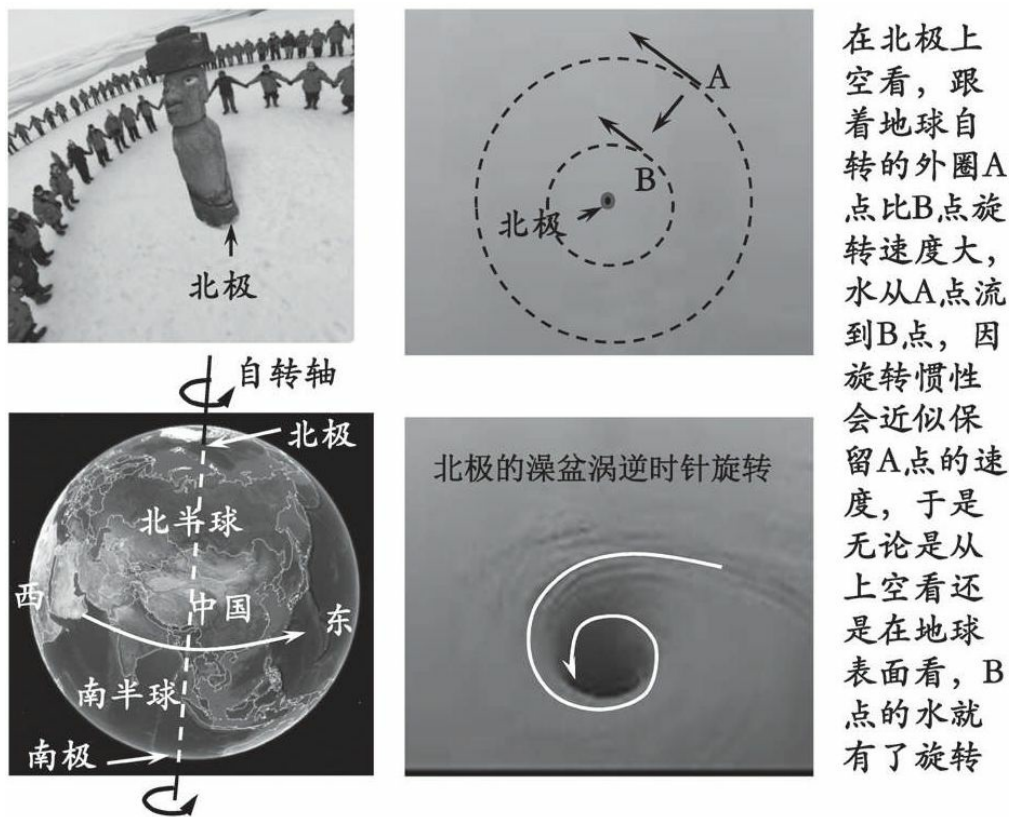


图2.24 地球自转导致澡盆涡的产生

在北极点放一个澡盆，让泄水孔正对着北极点。当然这不是个好地方，那里太冷，零下几十度，你得把水烧到一定温度才能做实验。但用于说明原理很恰当。在这里，水从周围的位置流向泄水孔时，原来随地球逆时针自转的旋转惯性要保留，于是越靠近泄水孔，旋转线速度越快，于是就能看到澡盆涡。在北半球其他地方，这种效果还存在，虽然不似北极点那样显得好理解，且效果弱一些，澡盆涡生成慢一些。对于南半球，道理类似。按此说法，在赤道上，就不应该有这种澡盆涡产生。

我们往往会用一种更难理解的道理去解释一个本来很浅显的道理。最常见的就是用科氏力效应解释澡盆涡的生成。

假设一只蚊子从圆桌的中心沿直线向圆桌外飞去。如果你站在旋转的桌子上跟着桌子转，再看这只蚊子，在你的视觉里它就不是走直线。于是，在旋转圆盘中的你，觉得蚊子受到了一个垂直于其运动方向的力，在你的视觉里它在偏离原有方向。1835年，法国科学家科里奥利（Coriolis）导出了针对运动问题在旋转坐标系看来不走直线的问题的力的数学表达式，发现这个力与物体运动速度成正比，与旋转角速度成正比，方向垂直于运动速度方向以及旋转轴方向构成的平面内。从20世纪开始，人们便把这个力称为科氏力。

那只蚊子本身没有受一个所谓的科氏力。所谓的科氏力，是在旋转物体中的你看到蚊子走曲线后，通过虚构这样一个力来解释这个蚊子为何相对于你在走曲线。本来是你跟着圆盘走弯道，蚊子不是。你以自己为评判标准，就觉得本来走直线的蚊子在走弯道。明明你受到了离心力的作用，却认为蚊子受了什么科氏力。

由于地球在绕自转轴由西向东旋转，因此人们往往把澡盆涡的产生归结于地球旋转导致的科氏力。其实这种解释看上去就是一种不一定符合逻辑的关联。

在地面上的人看蚊子在走直线，虽然在和桌子一起转的人看来是走曲线。因此蚊子是否走弯线，与观察者自己是否运动有关。澡盆涡不是这样，你外太空看澡盆涡在转，你在地球上跟着地球转，看那个澡盆涡还是在转。因此澡盆涡的旋转是绝对的，与观察者是否运动无关。

澡盆涡除了旋转，旋转中心的水位更低了，吸进去的空气柱还有噪声。恐怕你无法找到一个参照系，使你看起来噪声没了。因此澡盆涡的诱导因素似乎不能用科氏力来解释（虽然里面的一滴水的路径可以表示为与科氏力有关）。

澡盆涡为何会高速旋转（图2.25）

除了地球自转，容器几何不对称（即不是回旋体）、残余的带有某种旋转的流动和其他外力等，也可能诱发澡盆涡。

日常看到的澡盆涡，如果纯粹来源于地球自转，则可能不足以支撑那么高的转速。首先，澡盆涡每秒旋转的圈数，比地球自转每秒旋转的圈数，快成千上万倍。地球自转，每24小时转一圈，而我们能看到的澡盆涡，每秒至少一圈的量级，一天24小时接近十万圈。除此之外，拔掉塞子后，并不是立即形成澡盆涡，可能得等上较长的时间。在一些实验中，得等上十多分钟才出现澡盆涡。因此，地球自转、残余旋涡以及几何不对称等可能只是诱发因素，也许高速自转还有其他原因。

举例而言，家用或者实验用于产生澡盆涡的容器，半径也就半米量级，即使放在正北极，与地球一同自转带来的最外圈的旋转线速度也就每秒70微米左右。流到半径为0.7厘米左右的泄水口，旋转线速度按物理学原理也就扩大100倍左右，变成每秒700微米，即每秒不到1毫米，相当于每分钟转1圈左右，即每秒1/60圈。这么慢，都可能看不出来有旋转。

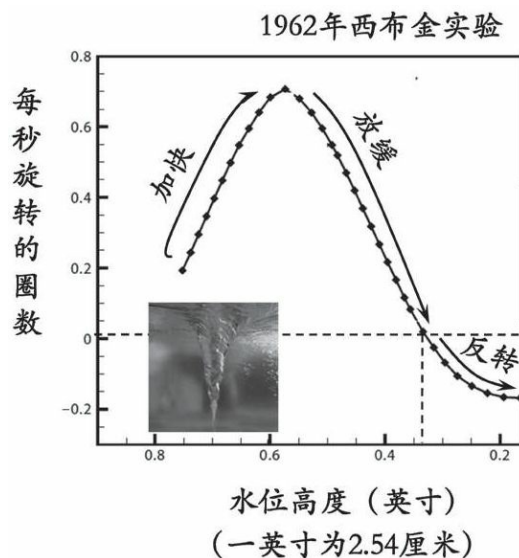


图2.25 澡盆涡转速先加快再放缓，最后反转

实际观测的旋转速度达到每秒1圈左右，因此，把地球自转的影响当作唯一的因素，不足以解释澡盆涡转速那么快。

因此，看似简单的问题却依然困扰着科学家，人们尚未完全弄明白旋转的本质原因。

旋转也是一种惯性。有了旋转，如果没有外力作用，旋转就不会停止。同理，让水转起来，也必须有外力作用。澡盆涡那么高的转速，尤其在涡心高速旋转，既然不完全来自于地球自转等诱导因素，有可能存在一些难以直观理解的，甚至有争议的内在机制。

1962年，西布金（Sibulkin）测量了一澡盆涡的旋转速度随水位高度下降的变化，发现澡盆涡刚开始转得很慢，每秒转 $1/5$ 圈，接着加速转速，最快可以达到每秒近1圈，随后转速开始下降，在水位下降到一定程度后，还出现反转现象。这说明，旋转速度与水位高度有关，不会完全来自于外力。水流旋转与容器底部的摩擦会搓出旋涡（虽然看上去不像旋转的旋涡），当水位降到一定程度后，这种搓出的旋涡区就占据主导，且被吞入泄水管时，引起的整体旋转与澡盆涡原有旋转方向相反。

容器下部的水贴着底部流入泄水孔时，要拐90度的弯，这可能会引起拐弯涡。如果是这样，这个旋涡就像一个环，套在下水口的入口位置。这样会部分地堵塞流水，不符合速度最快的演化原理。如果有旋转，即产生澡盆涡，那么就可以把这种拐弯涡转没，保证泄水速度最大化。也许这种用演化的角度来看待澡盆涡的出现更好理解。

5. 溪流、河道与海洋中的旋涡

澡盆涡是旋涡的一种。河边水下的小洞，抽水机入水口，水坝水管的入口，划水的船桨桨尖，都可能引起澡盆涡那样的旋涡。千里之堤溃于蚁穴，蚂蚁洞穴打通河堤后，浸水冲刷使洞越来越大，泄水过程如果出现旋转的澡盆涡，会加速将洞穴变大，导致溃堤。

曲水流觞（图2.26）

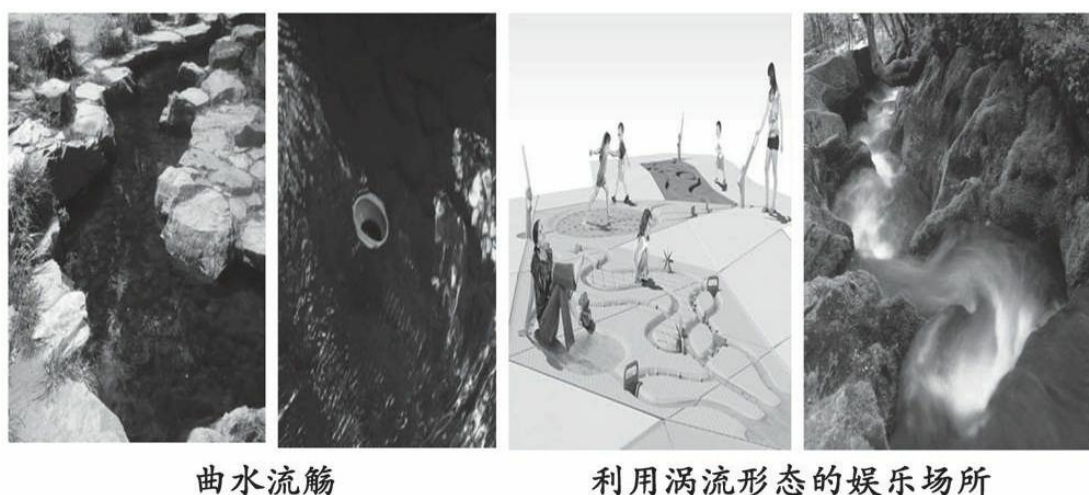


图2.26 曲水流觞与涡流旅行

溪流弯曲的地方，在水道拐弯急的地方也会出现拐弯涡。曲水流觞就是在弯曲的人造小溪流中，产生拐弯驻涡，即临时驻留在那里不移走的旋涡。顺流而下的酒杯（觞），如果正好落在旋涡位置，就可能驻留在那里一些时间，跟着一起旋转。

公元353年的农历三月初三，王羲之等人在兰亭溪流边席地而坐，他们将酒杯放在溪水的上游。溪流水道弯弯曲曲，容易在拐弯处产生旋涡，酒杯容易停留在那里片刻。顺流而下的酒杯，停在谁的面前打转转，谁就即兴赋诗。

有11人即兴各作出两首诗，15人各作出一首诗，剩下的16人则无法及时作出诗来。王羲之将这些诗收集起来并挥毫作序，写下了《兰亭集序》，其中关于曲水流觞的一段是这样的：

“永和九年，岁在癸丑，暮春之初，会于会稽山阴之兰亭，修禊事也。群贤毕至，少长咸集。此地有崇山峻岭，茂林修竹，又有清流激湍，映带左右，引以为流觞曲水，列坐其次。虽无丝竹管弦之盛，一觴一咏，亦足以畅叙幽情。是日也，天朗气清，惠风和畅，仰观宇宙之大，俯察品类之盛，所以游目骋怀，足以极视听之娱，信可乐也”。

中国目前兴起的曲水流觞艺术节，往往在有曲水流觞的溪流边以茶道和书画等形式陶冶情操。国外也有从水流得到启发的供儿童体验水流世界的涡流之旅。

旋涡的风采（图2.27）

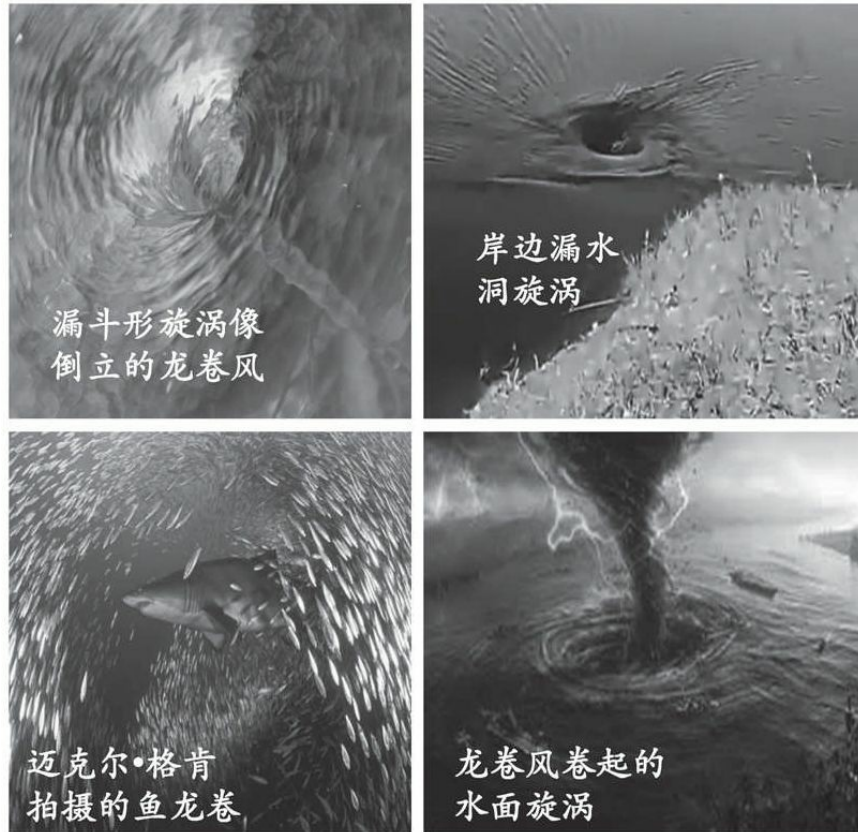


图2.27 水面与水中旋涡形态

旋涡有的很小，水黽腿蹬出的旋涡，不会比水黽大，直径也就是1厘米左右。水上龙卷风诱发的水中旋涡，会比龙卷风的底部还大。海洋中的旋涡半径可达100千米以上。

人们在大西洋南部发现的海洋旋涡，其特性像宇宙中旋转的黑洞，因此称为海洋旋涡黑洞。这种海洋黑洞导致的流动，加速了海洋污染物的扩散。之所以称为海洋黑洞，是因为这种旋涡的结构满足宇宙中黑洞的数学规律。如果你的小船在被卷进这样的旋涡，会螺旋般推往旋涡的中心，转速越来越快，很难逃生，就像黑洞有一个视界，进入视界后，连光线都逃不出去。

2016年9月21日，视觉中国报道了摄影师迈克尔·格肯（Michael

Gerken) 在美国北卡罗来纳州海域拍摄的鱼龙卷，可能是小鱼为躲避鲨鱼，往各个方向逃生时形成的图像。也许这只是一种视觉效果，看起来鱼群在围着鲨鱼旋转，好像鲨鱼滚水产生了漩涡一样。

漩涡是一种完美的圆形。只有这样，每一团水走的路径都一样，没有差异。大概只有这样，各部分的水团之间才能相安无事。等贵贱、均贫富。漩涡不可能无限小，因为与半径成反比的离心力不能无限大。转动不可能无限快，因为与旋转线速度平方成正比的离心力不能无限大。

2.3 轻抚水面生涟漪 强风破水摇巨浪

小物体轻轻触水，水面被压陷了，这还了得。幸亏有表面张力和重力在，它们使出女娲补天的本事，紧急修补。它们着急啊，使出的力量多那么一点点，就像荡秋千一样，把水位荡起来了。表面张力力量小啊，拖着短波浪使劲跑，重力力量大啊，挤着长波跑。不长不短的波呢？尤其那道1.73厘米的，就两不管了，就可怜地留在原地不怎么动啊。一圈一圈美丽动人的涟漪，就这样被表面张力和重力荡起来了，如同墨西哥人浪和蹲波队列。微风吹拂，到处都想泛起涟漪，都想是美丽的圆圈涟漪啊，但是叠加在一起，人哪能看得出来，只知道是猫爪印。糟糕，大风来了。表面张力波和重力波舞姿再好，也架不住狂风怒吼啊。但水不会屈服的，哪怕粉身碎骨，也要与大风搏斗。怒涛滚滚，巨浪滔天，那是水与风在厮杀。当然，水面规则波动时，不要想象得太复杂。我们在振动的世界中不是说了，往高度方向的一线水滴的集体运动，也像人的腿一样，近似地看也是左右的倒单摆运动和上下的弹簧运动的叠加。运动世界越简单越容易发生嘛，倒单摆与弹簧叠加运动再简单不过了。更复杂的连人都没法理解、连人腿都做不了的或者太累的运动，水才懒得做呢。一线水这样运动，就会拉扯和推挤边上的一线水滴做同样的运动，于是水面形态就是我们看到的波浪，往一个方向传播。

1. 蹲波队列 墨西哥人浪 幼儿园娃娃浪

旅游旺季，人潮涌动。这种描述显得很水感，在远处看，颇像微风吹拂下的水面小波浪，但一点也不整齐，也没有诗句“桑柘阴浓麦浪深”那种动感。如果有一种姿态协同，则可以形成动感十足的人浪，形态与我们看到的近岸海浪非常类似。

蹲波队列 墨西哥人浪（图2.28）

蹲波队列是一队人按照一定节奏原地运动产生的动感舞浪。蹲波队列可以用于检验或者激发团队的协同精神。看了蹲波队列，就能更形象地理解一些水面波浪了。

经过训练的士兵可以玩极其优美的蹲波队列。几十名士兵左右排成一排，面向正前方。排最左边的开始下蹲，当其中一个人看到左边的队友下蹲时，在0.25秒左右的反应时间内做出反应，也开始下蹲并站立。每人下蹲起立的姿势和速度尽量一样，且周期性重复。这样，每人的头顶连的曲线，看上去就像向左移动的波浪。虽然每一位士兵在原地蹲起，并没有向左移动，就像麦浪移动时，麦子留在原地一样。新的波浪不断从排头的士兵发出，向左传播。如果配以军乐，让军乐的节奏跟上波浪的节奏，那么场面可以十分壮观。

每个士兵在原地按顺序蹲起，形成波浪起伏的队形，波浪朝一个方向传播

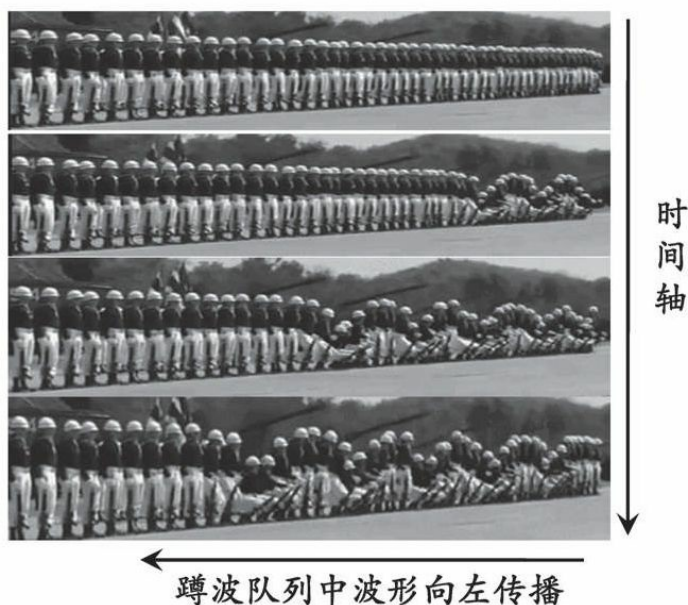


图2.28 士兵演练蹲波队列

电视上我们看到过墨西哥人浪（Mexican Wave），也称为人浪，或波浪舞，是观众席上的观众以排为单位依照顺序起立再坐下，获得可带动观众席气氛的波浪效果。这种由领头者激起的波浪舞，一般在欢呼进球时出现。

幼儿园娃娃浪

在幼儿园，可以将小孩站成一竖排，身体都朝着排最前面的小伙伴。每人都用双手搭在前面的小伙伴的肩膀上。排第一的小哥哥或小姐姐用一定的节奏慢慢蹲下慢慢起来，并不断重复。后面的在前面的带动下，一个接一个地重复同样的蹲起。协调得好，看起来像波浪运动。采用报数和蹲起相结合的方式，就能出现完美的协调。

也可以将小孩排成一个圆圈，由某位领舞。通过协调蹲起（或者摇摆身体）的节奏，使舞浪能在圆圈中和谐不间断地传播。如果人数较多，可以排成几个圆圈，甚至可以排出玫瑰花一样的队形。

如果追求动感，那么每人都唱同一首歌，但每人唱歌的起始时间和节奏，与自己的蹲起或摆动协调。经过适当练习和配合，场面极具表演感。这样可以让小孩锻炼出团队协作精神。

人浪的速度 身体的一点在画圆圆

无论是蹲波队列、墨西哥人浪还是娃娃浪，队员和观众并没有移动自己的位置，只是队形在向一个方向移动。就像我们看到风吹草地，花

草波浪形地倒下又起来，波形在传播而花草还留在原地一样。很容易设计出波浪的移动速度。

波峰与波谷之间的水面形状如同一条舞动的蛇，非常像正在弹奏的琴弦。这种波纹形状就是数学上的正弦曲线。队波中有最高点和最低点，高的称为浪峰，也称为波峰。低的称为浪底，也称为波谷。相邻两个浪峰之间的距离称为波长。

将相邻两人之间的身体中心轴之间的距离（比如说0.5米）除以每个人的响应时间即相邻两人之间重复同样动作的时间差（比如说0.25秒），就得到一个速度（比如说每秒2米）。这个速度就是波形的传播速度，也称为波速，即蹲波队浪从一侧向另一侧的传播速度。

每个人一蹲一起所花的时间（的平均值）也叫周期（比如说1秒），这个周期乘以波速（比如说每秒2米）得到的长度就是波长（比如说2米），即相邻两个波峰之间的直线距离，也是相邻两个波谷之间的直线距离。

如果相邻两个人之间的距离拉得足够远，且响应时间足够短，那么就可以得到极高速度的波浪传播。

蹲波队列的每个人在上下起伏，人并没有往波浪传播方向移动。如果玩蹲波队列的士兵，一边蹲起，一边左右摇摆身体，那么眼睛、鼻子、肚脐眼之类的就绕各自的原点在画圆圈。

小物体入水引起的涟漪，其中的水也是在这样画圆圈，而不是随波纹一起传播开。

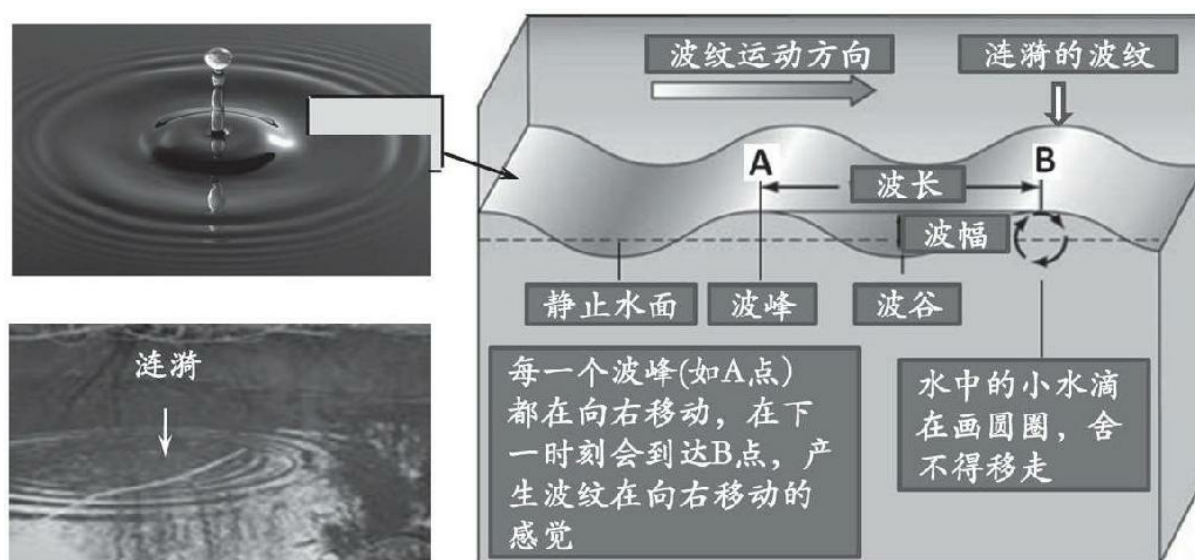
2. 不揭短的涟漪用美丽衬托客人

小虫子在水面跳来跳去，小雨滴落在水面，甚至小石块打水漂，都能激发一圈圈的涟漪。《诗经·魏风·伐檀》这样说：“坎坎伐檀兮，置之河之干兮，河水清且涟漪”。这可能是涟漪一词的来源。涟漪形态优美，有许多高贵的品质。

涟漪的形成（图2.29、图2.30）

有报道称，水龟在水上行走时激起了一对对旋涡。但我们在小池塘看到的是留下一串串圆形的涟漪。也许旋涡小得我们肉眼看不见。

设想在无风时，或者风速小于每秒23厘米的情况下，你放眼湖面，水面平静如镜。如果一片落叶或一滴雨水或一条小虫落在水面，或者你在打水漂时往水面扔一块小石片，你会看到原有平静的水面起了波纹，并且围绕落水点的波纹一圈一圈向外围传播。平静的水面受到扰动，产生的这种一圈一圈的波纹，称为涟漪。



来源：艾德·萨尔蒙《海洋科学》讲稿

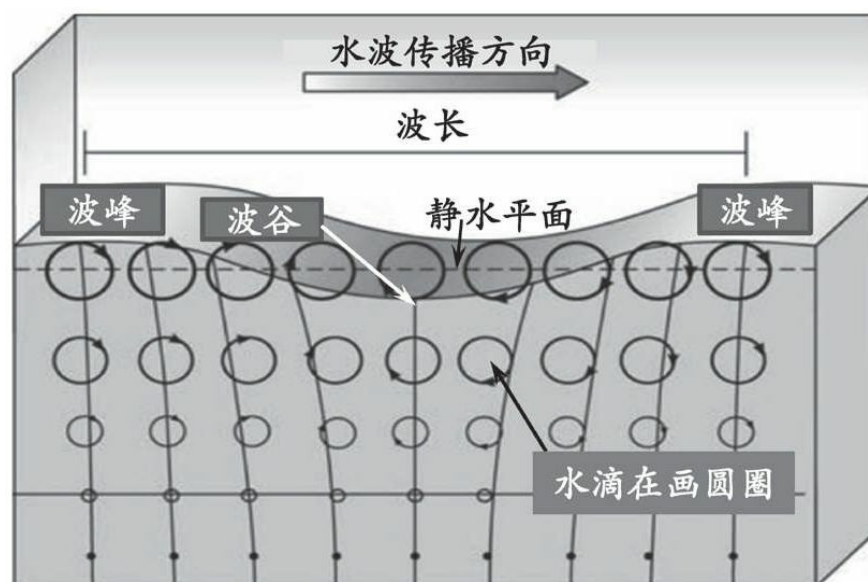
图2.29 涟漪圆圈型波纹的产生与扩展

与蹲波队列的人浪一样，涟漪中的波纹中的水面一高一低的，这种高低搭配的形态在移动。最高点也叫波峰，最低点叫波谷。相邻两个波峰或波谷之间的距离叫波长，即一段体现完整形状的水波的长度。可以简单地说，波长就是一个可见的圆圈波纹的宽度。每一个圆圈移动的速度叫波速，即波纹形态传播的速度。

高点多出的水来自于低点少掉的水。小物体落水时刻，沉下去一部分，当然把它周围的水挤高了些。小物体移走或沉下去后给水面留下空隙。空隙边上那些高出的水，试图来填平。这种填平的过程，会招致惯性力、表面张力和重力的共同起作用，互相拉来挤去，像荡秋千一样让水位振荡起来。

落水点高低水位的变化，通过表面张力拉扯周围的水位变化，也通过重力荡秋千一样的作用引起周围水位的变化，形成向外传播的波纹。

水波中的水滴，类似于蹲波队列蹲起时左右摆动的身体上眼睛、鼻子、肚脐眼之类的点，也在画圆圈或椭圆。水较深的情況下画圆，水很浅时画椭圆。



来源：艾德·萨尔蒙 《海洋科学》讲稿

图2.30 水滴在原地画圆圈，不离开原点

涟漪波纹爱圆形

小物体落水，不管小物体是什么形状，波纹居然都近似为圆圈，而不是小物体形状的放大。涟漪就是这样，不管触水的小物体长成怎么样，都会用美丽的圆环形涟漪来告诉别人小物体的来访，让别人觉得小物体的形状也很美丽。不揭短，不说坏话，把小物体美化，这就是涟漪。

原来，小物体的造访，只是激发了水位震荡。涟漪采用什么形态，完全是水位波动自身的规律起作用。给汽车设计师奖励、激励和待遇，设计师就设计出自己熟悉的作品如未来汽车，而不是设计出奖状和人民币一样形状的汽车。

如果小物体推开水位形成的初始水位形态不是圆形，而是看上去凹

凹凸凸的，那么凸出的就会被表面张力拉住，跑得慢，凹进的就会被表面张力拉扯，跑得快，直到出现圆圈形状。重力的影响也类似，虽然理解起来没有那么直截了当。

3. 不插队不打架的涟漪波纹

小涟漪尽是优点。除了不揭短，还不插队。我们可以跑到附近的湖面、池塘或小水沟去扔小东西生成涟漪，目测涟漪波纹的波长，以及涟漪移动速度大概是多大。我们会发现，涟漪波纹的波长有一些奇特现象，要么越在外面的波纹波长越大，要么反过来，越在内圈的波长越大。

内圈与外圈的波长（图2.31）

你可以试试不同大小的物体激发的涟漪。比如说用直径半厘米左右的小水滴去滴水面，激发涟漪，或者用巴掌大的石块打水漂激发涟漪。检查是否出现这样的情况：小水滴滴出的涟漪，越往外的波纹，波长一般越小。而大的水漂石片激发的涟漪，越往外的波纹，波长一般越大。

物体触水，激发的涟漪可能出现不同波长的波纹。各种波长的大小与落水物体的尺寸当然会有关系。物体越小，涟漪各波纹的波长越小。物体越大，涟漪中各波长越大。这与直觉是一致的。

不同波长的波纹，受到的表面张力和重力影响的程度显然不一样。于是，不同波长的波受到的力的大小不一样，移动速度当然会与波长有关。

什么波会留在内圈？肯定是波速最小的留在内圈。如果仔细去测量，会发现留在内圈的波纹的波长，大致不到2厘米（实际上是1.73厘米左右，具体大小与温度有点关系）。这说明，这个波长的波传播速度最慢，且可以测得（科学家用理论也可以推导出），这个最慢的波的波速大致等于每秒23厘米。往后我们会理解，为何有一个速度最慢的波。

波长比1.73厘米小的波纹，以及比1.73厘米长的波纹，都比这个波长的波快，于是都跑到外圈了。波长小于1.73厘米时，波长越小的波纹传播速度越快。反过来，如果波长都比1.73厘米大，那么波长越大的波纹，传播速度越快。无论如何，留在内圈的，传播速度最慢。

外圈的涟漪波纹波长小 外圈的涟漪波纹波长大



图2.31 不同大小物体落水激发的涟漪

对于比1.73厘米小得多的小水滴产生的涟漪，波长一般都比1.73厘米小，这样的波，波长越短的跑得越快，于是越外圈的波长越小。对于比1.73厘米大得多的水漂激起的涟漪，会有一些波长比1.73厘米大，因此越大的波长的波，越跑到最外圈。

涟漪的叠加（图2.32）

当然，对于同一次产生的涟漪，依据物体形状、大小以及落水方向和力度不同，也有可能产生波长既有比1.73厘米长的也有比1.73厘米短的多个波纹，这些波都会比1.73厘米的波跑得快。也有可能入水物体在那里震荡几下，激发几组不同的涟漪。每一组涟漪都有几串波纹。不同组的波纹相隔一定的距离。一个组也可以称为一个群。在每个群中，波长随半径单调增加或减小。

涟漪还有一个优点，就是不打架。两个相邻涟漪的波纹交汇时，各走各的，相互之间不拦路，不干扰。温柔善良的涟漪，当然不会相互动粗。

那么，为何有一个波长，对应的波速最小？这就是重力和表面张力的力量对不同波长有差异的结果。

小水滴落水后震荡，激发几组涟漪，每组包含几道波纹

两个相邻涟漪的干涉

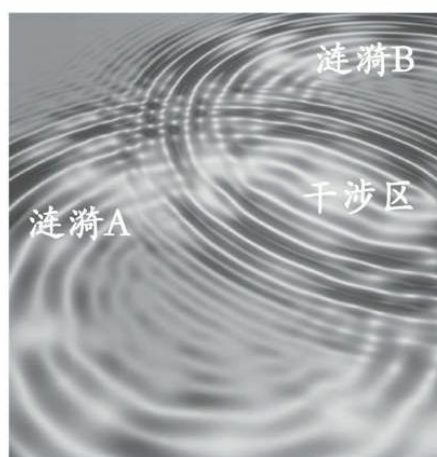
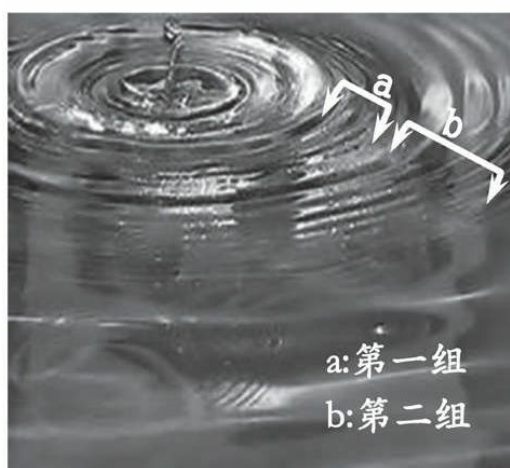


图2.32 一点持续产生涟漪以及两个涟漪的干涉

4. 波速也会论短长 表面张力与重力的分工

表面张力力量小，当然喜欢驱动波长小的波，重力力量大，只看得起波长长的波。一个爱小，一个恋大，那种不大不小的波长，当然就被晾在一边啦。那道1.73厘米的波长也不生气，你们恋大恋小，我就恋家吧。再说，每秒23厘米的速度，也够在周围转转了。水面有波动时，重力将高水位的水滴往下拉，表面张力则总是从与水面平行的方向拉水面。表面张力力量小一些，正好够拉扯波长小的波。重力力量大一些，不好意思挑小波长的波，因此主要推动大波长的波。也可以说，对于小波长，表面张力唱主角。对于大波长，重力唱主角。于是，波长很小和很大时，表面张力和重力分别胜任作为驱动它们的角色。

表面张力波的波速（图2.33）

波纹使水的表面弯曲了。一段波长的波的两端被表面张力拉扯，波就振动了。一页弯曲的纸张，我们从两侧拉扯，纸会被拉平。同理，表面张力也会把波谷拉高波峰拉低。这种拉扯也使波中水滴获得了速度，被拉平后，这种速度就成了惯性，驱动波纹向另外一个方向震荡。

一段波包含了一定的水量，里面的水都得震荡起来，这种震荡就是表面张力在来回使劲的结果，如同在荡秋千。

波长越大，水量就越多，表面张力只在两端拉扯，当然就费劲。就像你从端点拖动一个物体，只有那么大的劲，当然物体越长，拖动越困难。于是，由表面张力决定的水波，波长越大，传播速度越慢。也就是说，表面张力单独起作用时，引起的涟漪波纹，传播速度随波长增加而单调下降。

表面张力拉动一个波长的波走一个波长的距离，做的功等于表面张

力乘以这个距离，做功使这段波获得动能。如果波形给定（波幅比波长是固定值），这段波包含的水的质量当然正比于波长的平方。由于动能是质量乘以速度平方（的一半），因此这样获得的速度就与波长的根号成反比。表面张力波是一种振荡型的做功与动能变化，但道理和刚才的分析是一致的，从而波速也与波长的根号成反比。

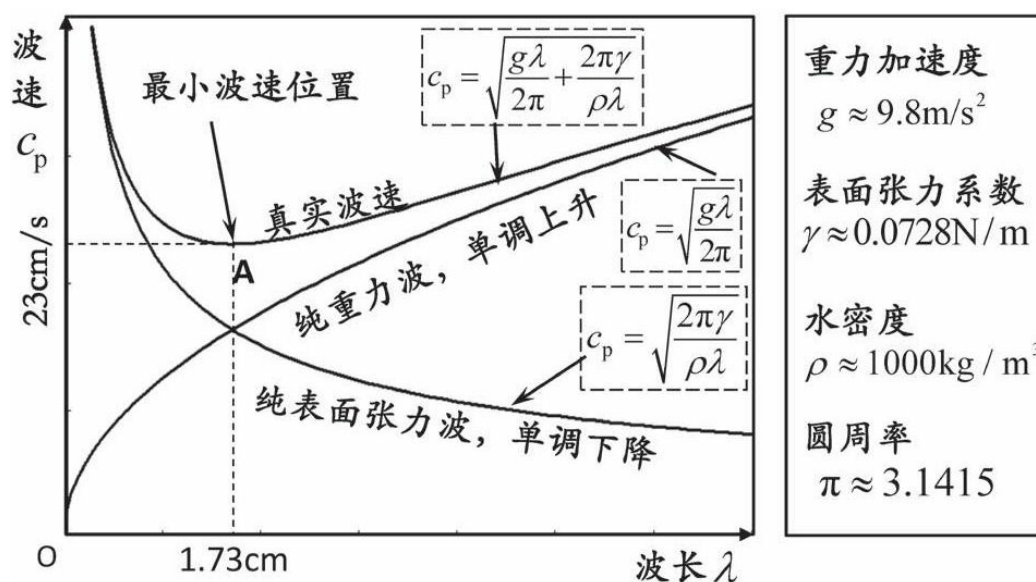


图2.33 波长为1.73厘米时波速为最小（每秒23厘米）

重力波的波速

水面有波动时，重力将高水位的水滴往下拉，拉出惯性来，于是拉平后凭惯性进一步往下震荡。这一震，就震得比周围水位还低了（和荡秋千道理一样）。周围显得更高一点的水位压出的高水压，迫使低位的水又往上震荡。于是就来回震荡起来。由于是从落水点激发的震荡，因此多余的力量就驱使波浪向外传播。

重力不像表面张力，不是在一段波的两端拉，而是往下压往上举。

一段波包含的水量显然与波长成正比。由于重力正比于水量，因此波长越大，用于改变形状（从而导致波纹传播）的重力就越大，作用就越强，波速就越大。于是，重力决定的波速随波长增大单调上升，即波长越大，波速越快。当然，这样的解释还显得很抽象。

为了进一步理解，可以试图与单摆的振动比较一下，假设你更理解单摆的话。单摆的摆线越长，摆动周期越大（也可以说摆动频率越低），周期与摆线长度的根号成正比。水波中的重力波的震荡周期（即波峰波谷一下一上完成一次震荡需要的时间周期）也是如此，必然与波长（与单摆摆动幅度从而摆线长度类似）的根号成正比。你用波长除以这个周期就得到波速。波长除以波长的根号，就剩下根号波长了。于是，重力波的波速，与波长根号成正比。

也可以用另外一种方式理解。我们知道在一定高度的物体有重力势能。让它掉落，获得的速度就与重力势能的减少有关，获得的动能（与速度平方成正比）正比于下落的高度，因此获得的速度是这个高度的根号。水波中水滴的上下震荡也是如此，水滴上下震荡的速度就正比于偏离原始水平面的高度的根号。水波形态向外运动的速度（波速），必然与这个上下震荡的速度的大小（某个绝对值的平均）成正比，而对于形态一样的波，振幅必然与波长成正比。于是，不难理解，重力波的波速与波长的根号成正比。

表面张力和重力的综合作用 极值（图2.33）

实际上，表面张力和重力同时在起作用。那么是纯表面张力波的波速和纯重力波的波速直接算术相加得到波速吗？不是，它们也要使用勾股定理。勾股定理说，直角三角形的弦的平方，等于勾的平方加股的平

方。于是，弦长等于勾的平方加股的平方得到的值开根号。波速也是如此，涟漪水波的速度是纯表面波波速的平方与纯重力波波速的平方相加后再开根号得到的速度。波速之所以符合这种平方叠加规律，因为其遵循能量守恒原理，即能量可以直接线性叠加得到总能量。表面张力和重力使水波中的水滴有了能量，两者相加就是总能量。而能量正比于速度的平方，因此能量叠加就是速度的平方叠加。

现在是表面张力和重力同时起作用，分别给出的波速与波长之间满足相反的规律，一起发生作用时，就看哪个作用大。波长特别小时，表面张力波起主导作用，波长特别大时，重力波起主导作用。一个团队的事情，小事情如卫生、烧茶端水、收发信件等，是级别较低的人做得好，大事情如产品设计、融资和重大决策等则是级别较高的起作用。对于中不溜的事情，低级别的和高级别的作用都不大。因此，中不溜的事情往往没人管，效率低。

同理，当波长不大不小时，表面张力玩不了大的波长，重力又瞧不起小一些的波长。最终结果是，好像都不怎么管，使波速较慢。不明真相的，会用一个和尚挑水吃、两个和尚抬水吃、三个和尚没水吃来解释为何表面张力和重力都应该起作用的波长，波速反而小了。

既然不大不小的波长的波会更慢，那么必然有一波长的波，其波速最小。这就是前面提到的在波长近似为1.73厘米时，波速最小（约23厘米/秒）的原因。

总结而言，波速随波长变化的曲线，对于表面张力波和重力波，分别单调下降和单调上升，两种共同作用时有最小波速，数学上叫存在极值。这种极值现象，在水波以外的其他方面也较普遍。在遇到类似的由影响趋势相反的两种因素起作用的现象时，就可以思考是否存在极值。

有时是在某中间位置取极小值，几种因素在中间区域都鞭长莫及的结果，一个和尚挑水吃，一加一小于二，甚至小于一，就是这个道理。有时是取极大值，好像一加一大于二，三个臭皮匠顶个诸葛亮。还有一种取极值的机制，如同环境污染的演化、流行病住院病人的演化以及汉字数目随笔画数目的变化，都在某时刻或某位置取极大值。这是因为，增长因素使数目少时快速增长，但衰减因素则在数目多时变得异常强劲，使数目无法突破某一极值。这就是少的时候没人管，多了后招致嫉妒的结果。

5. 各种风速下的风浪

小物体落水，或者微风吹拂，水面会很温柔，尤其生成美丽的涟漪。可是，如果风太大，水面就招架不住。你可以说，是大风掀起了大浪。也可以说，水面升起大浪来与大风搏斗，削弱大风的力量，使大风无力在更远的地方撒野。当然，你也可以认为风与水在一起狂欢。狂风怒吼时，水中巨浪滔天、惊涛拍岸。

风吹大海，水面会被风吹起涟漪或水浪。或波光粼粼，或惊涛拍岸。如果超过8级大风，那么连船都不敢出航了。

打破水面的平静 零级风的风速

如果风速在每秒23厘米以下（零级风，也有的把每秒30厘米以下的风称为零级风），那么水面会平静如镜，没有任何浪花，如同我们水缸中的水面一样。

由于水面表面张力的抵抗，低于每秒23厘米的风，无法将水面吹出

波纹。我们可能会诧异，每秒23厘米是怎么来的。

原来，已经知道，涟漪的波纹最小传播速度是每秒23厘米。因此，如果风以每秒23厘米以下吹水面，即使在局部产生波纹，那么这种波纹也会以每秒23厘米以上的速度向四周散开。风刚刚引起一点点水面扰动，就传播走了，这是风跑不过涟漪的结果。

微风中的涟漪叠加成猫爪印（图2.34）

如果风速高于每秒23厘米，那么就可以缠住那些速度更慢的波纹，不断吹拂，将其放大。于是，如果风速在每秒23厘米以上，水面的平静就破坏了，破坏程度与风速有关。

风速在每秒0.3~1.5米之间时，称为一级风。此时，在水面会产生波光粼粼的涟漪，没有明显波峰，浪高在0.2米以下。

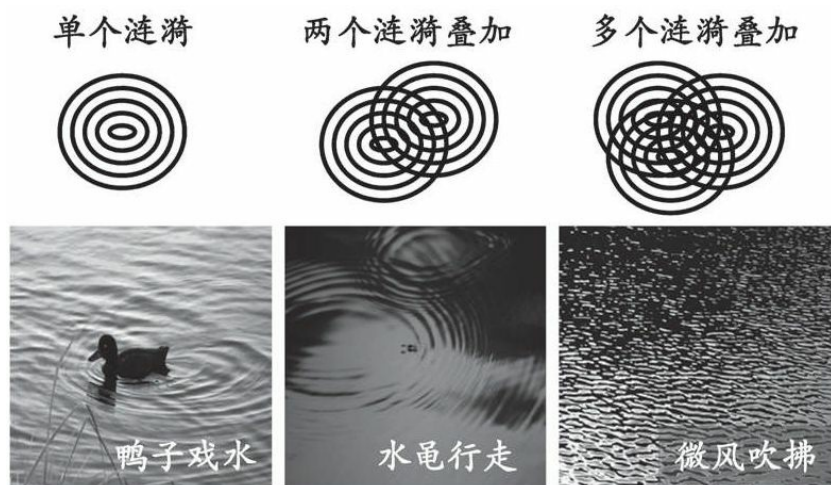


图2.34 多个涟漪叠加可形成猫爪印式的波纹

学术界所指的涟漪（ripple）是小物体落水或者微风吹拂水面产生

的波纹。如果你家有宠物猫，那么可以看看猫爪，至少可以看看猫爪印。你会发现，猫爪印与微风吹拂水面产生的涟漪波纹很相似。因此，英文中把微风导致的涟漪也称为猫爪（cat's paw）。小物体落水产生的涟漪是圆圈型波纹。两个相邻的涟漪波纹叠加时，就会出现波纹交织的小格格，如同渔网一样。

微风吹拂时，风中的大气脉动可以想象成是无数小物体落水，它们分别激发圆圈型涟漪。这些一个挨着一个的涟漪圈圈，叠加后就成了猫爪印似的涟漪了。

风速的增加（图2.35、图2.36）

风速超过一级后，水面波浪肯定比涟漪更强。浪高随风速增加而增加，是因为风给水注入了能量，风速越大，浪当然越高。

风速在每秒1.6~3.3米的二级风能在海面吹起浪高在0.2~0.5米之间的鱼鳞状波纹，这种波纹不会碎裂。

到了风速为每秒3.4~5.5米的三级风，海面会产生浪高在0.5~1米之间的带有白盖的鳞波，波峰开始碎裂。

风速在每秒5.5~7.9米的四级风，在海面引起短波波浪，波峰碎裂，频繁出现白盖，浪高在1.0~2.0米之间，有的超过人的身高。

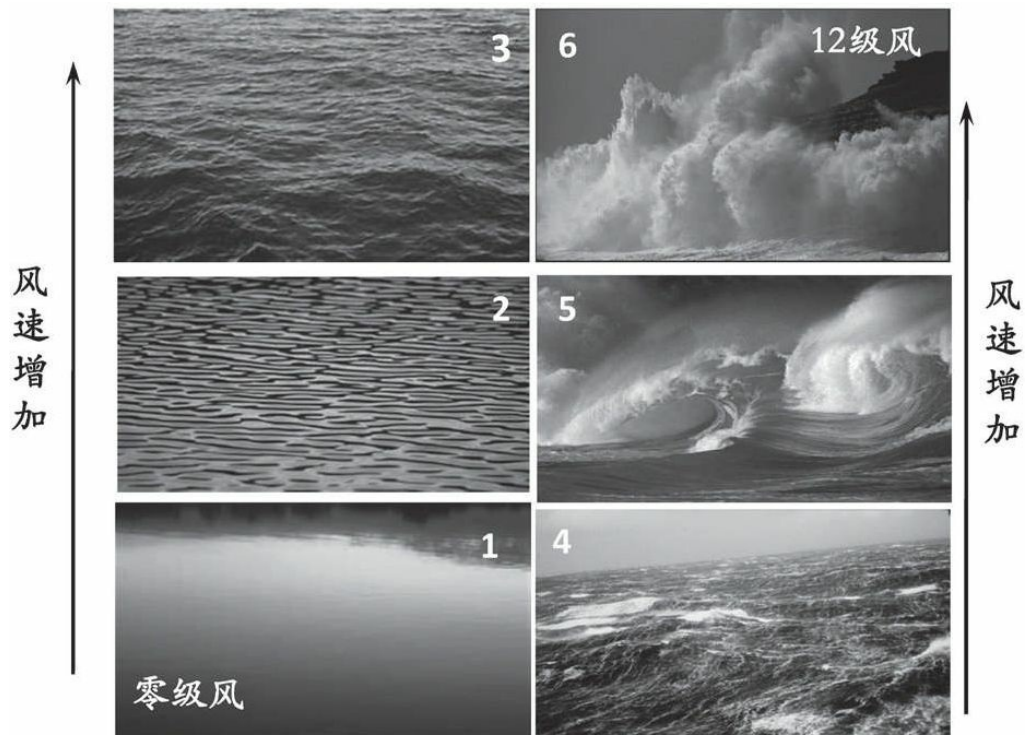


图2.35 不同风速下的水面波浪

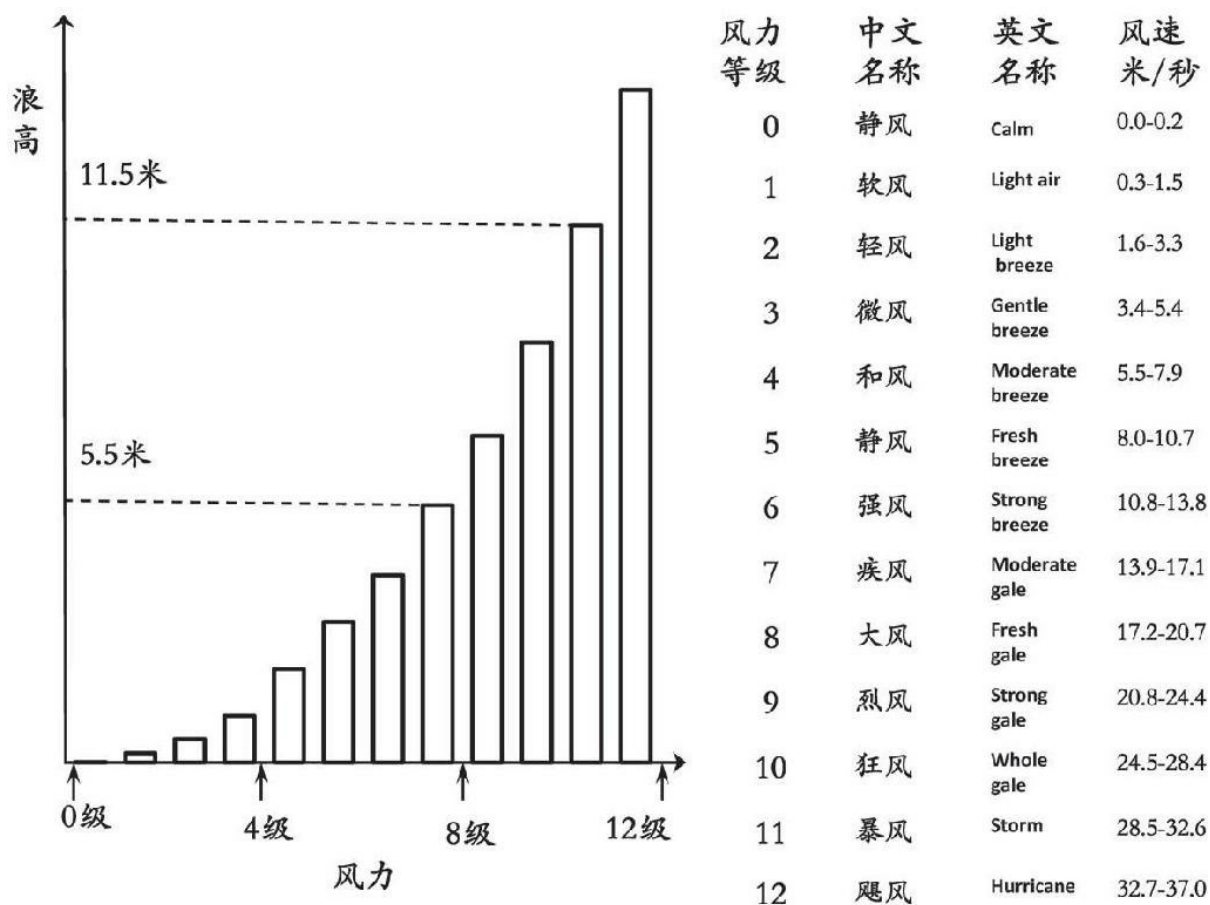


图2.36 浪高与风力等级（不同来源取值范围略有差异）

风速在每秒8~10.7米的五级风，让海面产生中尺度的波浪，上面铺满白盖，并且有少量水雾飞溅，浪高在2.0~3.0米。六级风的风速在每秒10.8~13.8米，超过了飞人博尔特的速度。海面产生长条形波浪，出现白色泡沫状波峰，一些水雾携带空气，浪高在3.0~4.0米之间。七级风的风速在每秒13.9~17.1米，海面产生沙丘一样的波浪，一些泡沫组成随风飘舞的条带，出现中等数量的水雾，浪高达到4~5.5米，超过了一层楼的高度。

大风的破坏力 惊涛骇浪（图2.35、图2.36）

八级风就是大风了，风速在每秒17.2~20.7米，每小时超过60千米。海面有5.5~7.5米高的风浪，浪花和成型的泡沫带会顺着风向，出现大量水雾，港口的船只一般不会出海。

九级风的风速高达每秒20.8~24.4米。海面有7~10米的高水浪，波峰有时卷曲，高密度的泡沫顺着风向，大量水雾会降低能见度。

十级风的风速到了惊人的每秒24.5~28.4米。海面有9~12.5米高的水浪，波峰倒挂，大块的泡沫使水面变成白色，波浪翻滚，惊涛拍岸，大量水雾降低能见度。

十一级风的风速在每秒28.5~32.6米，相当于每小时100千米左右，几乎到了高速公路每小时110~120千米的限速。海面水浪11.5~16米，激起大块泡沫覆盖水面，巨量水雾降低能见度。

我们听到十二级台风就非常惊恐，风速达每秒32.7米以上。海面巨浪滔天，浪高超过14米，比四层楼还高。水面全是白色的泡沫水雾，空气充满漂移的水雾，极低的能见度。

浪高的增长水雾的形成

风速增加，浪高也增加。单看数字显得很平淡，但如果放在图上看，就不一样。让风速翻倍，会发现浪高不是翻一倍，而是更多倍。这种翻倍数目不一样，在数学上叫非线性。比如说，8级风相比于4级风，风速增加了1.8倍，而浪高增加了3.3倍。

为何不是同样的翻倍数目？原来，风浪的形成肯定与大风注入的能量有关。这里当然是动能起作用。动能正比于风速的平方。因此，风力

翻倍，动能翻两番，于是浪高翻的倍速就比风速翻的倍数多。当然，浪高并不是恰好是翻两番，因为激发浪涛时，一部分能量也转变成水的速度了，并不完全用于抬高水浪。当然，还有一部分能量被摩擦和碰撞损失掉了。

两个小泥团一撞，会被撞碎。巨浪高速撞击，也会撞击出细碎的水花来。大风中的小尺度乱流，也会挤出一些小尺度液团，尺度小了后，表面张力一包裹，就剥离出小液滴了。小液滴多了后，看上去就是白雾。

如果大风吹出了一块块片状液体，那么就像风中的旌旗，会舞动，甚至卷起成葱油饼似的形状。毛泽东的诗句中，就有红旗漫卷西风。连红旗都能被风卷起来，何况柔顺得多的片状液体。

这里描述的水浪似乎不怎么有形，但浪涛并不全是乱的，否则谁还敢玩冲浪运动。比如，在海岸边我们能看到一些长条状的波浪，一浪一浪地过来拍打海岸。哪怕当地没风，偶尔也能看到。它们从何而来？

2.4 奇妙无比的水面世界

涟漪因为小、形态简单而美丽。水面世界不仅仅只有涟漪，还有形态万变的风浪、水位高低振荡的湖震、波长堪比地球半径的潮汐、钱塘江呼啸而过的大潮、在海洋中隐秘不可探测但传到海岸后威力无穷的海啸、从船边碎浪长出的形态简洁的船头波和船尾波。无风不起浪，无风三尺浪的秘密就会在这里出现！水面的一些形态和规律，也和大气中飞行发生的一些现象很类似。例如，船边有V字形波浪，候鸟排V字形队列。

1. 无风不起（风）浪 无风三尺浪（涌）

无风不起浪，即本地产生的风浪来自于本地的风。无风三尺浪，即其他地方的风浪变成了涌浪，传到这里来了。风平浪静，是指风小时，水面就平静如镜。微风吹起，水面波光粼粼，产生了涟漪，但涟漪波纹的传播好像与风向没有关系。风再大一点，风浪起来了，风越大，风浪越高，风指向哪，浪打到哪，直至惊涛拍岸，风停则浪止。可见，由风激发的水面波浪的大小甚至形态，与风的速度和强度有关。在大海边上，有时突然从远处冒出长条形波浪，越到岸边浪越高，速度越慢，甚至非得横着撞击海岸。原来，浪速与水的深浅也有关系。

涟漪，风与水的缠绵（图2.37）

如果风速小于每秒0.23米，水面会平静如镜，除非有昆虫在水面爬行、有小鱼在水面戏水、有船舶游弋。如果风速稍微高一点，那么由于

风中的空气与水面的摩擦作用，水面会产生涟漪，在太阳照射下波光粼粼。但这种微风产生的涟漪，朝各个方向传播，而不是单一地朝风的方向传播。这与水毬滑跳激发的涟漪相似，只是不再是波纹圈圈形式，而是一个一个小水丘在交替出现。涟漪波纹的传播朝各个方向，与风向无关。如同幼儿，还不知道听话。涟漪是在风的作用下即时产生的，风停止后，很快就没了，表面张力把它们消耗掉了。

看来，涟漪是风与水谱写的
《我侬词》（元代管道昇）：你侬我侬，忒煞情多，情多处，热如火。把一块泥，捏一个你，塑一个我，将咱两个一起打破，用水调和，再捏一个你，塑一个我，我泥中有你，你泥中有我。与你生同一个衾，死同一个椁。

当然，你可以说，微风在水面各处激起圆圈涟漪，不同位置的圆圈波纹叠加在一起，就成了猫爪印了。

无风不起浪 风浪 (图2.37)

如果风速进一步增大，且朝一个方向吹数小时，那么就会产生清晰可见的风浪（wind waves），波



涟漪像猫脚印



风浪条纹不规则



涌浪条纹规则

长在几十厘米到一百多米之间。这  图2.37 涟漪、风浪和涌浪

种风浪沿着风的方向，风越强，风浪越高。风往哪里吹，哪里产生风浪。风停止后，风浪就停止了。如同青少年，知道听话了。无风不起浪，指的就是这种现象。风改方向，风浪也改方向。

风浪的强度、波长和波速均与风的强度、风吹的时间以及吹的距离有关。如果是规则的长条形波浪，则波长往往是浪高的七倍以上，浪高如果高于波长的 $1/7$ ，那么高出的部分就可能碎掉溅出白花。

无风三尺浪 涌浪（图2.37）

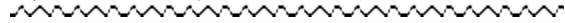
风浪来自于大风不断给海水注入能量。风在哪里，哪里就有风浪。无论吹多久浪都是那么高，且风浪不是那么规则，奇形怪状的。难道不断吹风注入的能量，除了维持浪高，多余的全损失掉了？不管什么形状的小物体落水，虽然被搅和的初始形态也一定与小物体形状有关，但最终激荡出美丽有形的涟漪向外扩散。同理，大风持续吹拂，虽然产生形状不规则的大浪，但深水被持续激荡，也会产生形状规则的有一定波速的水波，向远处传播。这就是涌浪，属于长条形波浪。

风很强且长时间吹水面，比如说吹10多个小时，那么在风浪中就会产生涌浪（swell）。涌浪是强风持续给风浪注入能量后出现的清晰可见的具有长条形波纹的水波，按深水波的自身规律演化和维持形态。就是说，涌浪虽然来自于风浪，但有了足够的能量和自己的特征，有特定的波长和波速，与风没什么关系了。成年人学成后，有了自己的主意和能力，会按自己意志行事了，往往早离开了学校。

风停了，涌浪还在。风转向，涌浪不转。涌浪的形态比风浪更单一，只保留了风浪中更有形、更定向的水波。涌浪可以传播到很远的甚至没有风的地方。无风三尺浪指的就是从别处传来的涌浪。

涌浪的波长与引起涌浪的气象有关系，一般很难超过150米。我们在海岸看到的大部分波浪，其实来源于远海的涌浪。

水深与波浪（图2.38）



涟漪的波长与波幅，与水深相比，太小了。因此，只在水面戏水的涟漪，感觉不到水底的存在。

如果浪的波长很大，或者相比于波长，水很浅，那么重力引起的压力波就能把水面波动信息传送到水底，水底反射压力波，反过来会影响水面波动。

水的深度如果超过波长的一半，那么这种水底效应就很弱，这样的水波称为深水波，表面张力和重力均起作用。

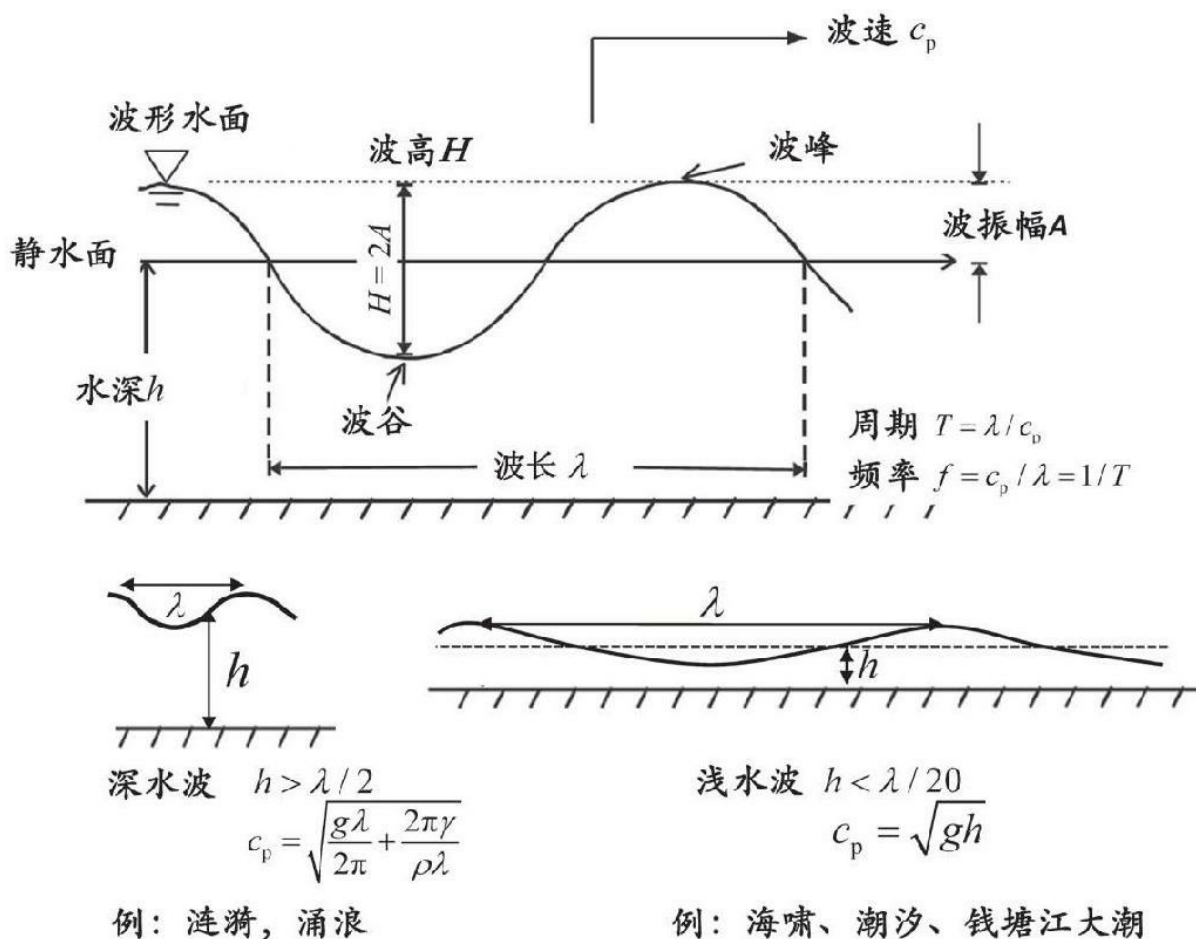


图2.38 深水波与浅水波

如果水深只有波长的1/20或以下，就是浅水波，重力会决定水波的形态，完全不让表面张力起作用。水浅时，高度方向能形成的水压与长度方向的水量相比显得少多了，因此不容易驱动水波运动，因此水越浅，波速越小。由于这种波的传播是来源于与水深成比例的重力势能的释放与聚集，因此动能正比于重力加速度和水深的乘积，速度是其乘积开根号。波速也与这个速度有关，于是波速是重力加速度和水深的乘积的根号。

介于两者之间的水深，表面张力和重力会同时起作用，那时就很复杂了。

涌浪靠近岸边时（图2.39）

涌浪从远海接近海岸时，由于水深越来越浅，会慢慢变成浅水波。水越来越浅，速度也越来越低，因为水越浅，浅水波的波速于是越小。

越靠近岸边的波速越小，因此本来与海岸线不平行的远处水波，近岸时也近似与海岸平行，且水浪越来越高，超过波长 $1/7$ 从而破碎成白雾。

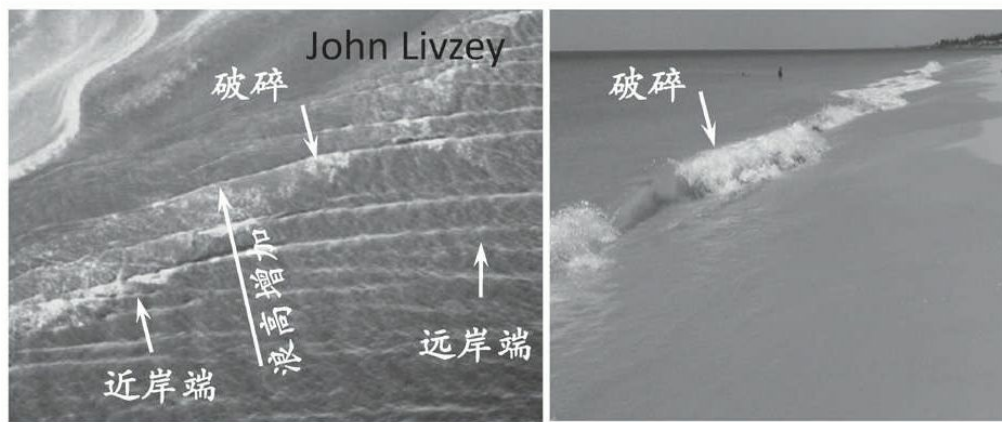


图2.39 波浪近岸时与岸边近似平行

那么能量到哪里去了？涌浪包含的能量还与水深有关。到了浅水的岸边，水深浅了，波速慢了，因此，多余的能量就会推高水浪。于是，越靠近岸边，浪显得越来越高，因此有时感觉在岸边突然冒出涌浪，因为在深海区涌浪可能隐藏在水下。

我们在海边观察近水波浪，会发现不管远处的波浪是否与海岸线平行，到了岸边，一般会近似与海岸线平行。原来，即使首先不平行，先接近海岸的一端，由于提早遇到浅水从而波速减慢，远的一端则以更高的速度接近，最终就近似与海岸线平行了。

近岸波的浪高与波长相比，很快超过七分之一波长，导致破碎。因此，近岸波很容易出现碎花。

2. 游泳池的震荡 湖震 反射与叠加

将盛有半杯茶的杯子摇动几下放在桌面上，杯中的茶就在晃动。浴缸和鱼缸中的水也可能晃动起来。晃动时，一侧的水位升高，另一侧的水位降低，如此交替反复。这种晃动当然也是一种振动类型。游泳池的水也可能晃动起来，主要在放水过程容易引起晃动。

在半封闭或者全封闭的港湾或湖中，强风和大气压力的迅速改变，会将水从一侧推向另一侧，使另一侧的水位更高。风停止后，水从水位高的一侧向水位低的一侧运动。于是，水位就来回震荡，可持续数小时甚至几天。这种现象称为湖震。湖震是一种驻波，即波纹看上去在原地震荡。

水从一侧推向另一侧，在岸边会反射回去，改变水波的传播方向。在游泳池、蓄水池、港湾、内湖、海港、海洋等水的中央如果有水波形成，那么水波向岸边传播，碰到岸会反射，这种被反射的水波可称为入射水波，反射回来的称为反射水波。反射水波朝入射水波相反方向走，就像乒乓球从墙面反射回来一样。反射水波与持续而来的入射水波会叠加在一起，形成叠加后的形态。

设想一道水波的波峰以波速向右移动，一道具有同样波长和波幅的水波即反射后的水波向左移动，速度一样、方向相反。如果在某点，右行水波的波峰正好与左行水波的波谷重合，那么该点的水表面就没有高度和深度了。如果右行水波的波峰正好与左行水波的波峰重合，那么该点的水位就叠加成两倍的波峰高度。可以证明，相同波长和相同速度但

方向相反的两道水波叠加后，会形成一道驻波。这种驻波的波纹形状在上下震荡，看不出波形水平移动了。

由于湖震（图2.40）可以看成是两个方向的浅水波的叠加，每一道浅水波波速是重力加速度与水深的乘积的根号，因此来回一次震荡的周期就是两倍的水池长度除以这个浅水波波速。这个关系称为梅里安（Merian）公式。

假设游泳池的长度为50米，水深为2米。那么按梅里安公式，湖震的震荡周期是22秒左右。如果是在长度为5千米、平均水深5米的湖中产生湖震，那么震荡周期就是23分钟左右。

伊利湖（Lake Erie）是北美五大湖的第四大湖。其北面是加拿大，其他面为美国，西南至东北长388千米，最宽处92千米，平均深度18米，最深64米。如果用Merian公式计算长度方向形成的湖震震荡周期，那么周期是16小时左右。如果按宽度方向，周期是4小时左右。据说这里是最容易在强风作用下形成湖震的地方。湖震浪高可达数米。一些大型湖的湖震周期在4小时到7小时之间，非常接近海洋中潮汐的周期，因此人们往往把湖震误当成潮汐。

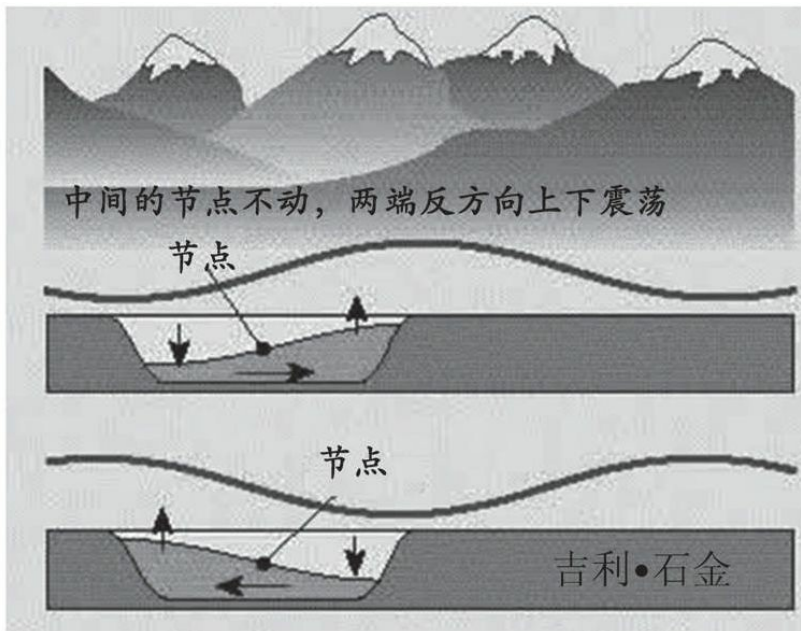


图2.40 日内瓦湖的湖震是具有半个波长的驻波

3. 海洋潮汐 钱塘江大潮 海啸

月球很遥远，我们算我们的重量时，恐怕不会计算月球引力的贡献。但是月球引力因月球运动而随时间变化，会对海水局部受到的引力造成变化，这足以引发了海洋潮汐，海洋潮汐传到内河，形成大潮。最著名的就是钱塘江大潮。除此之外，地震、冰山滑落等还可能引起海啸。

海洋潮汐

地球是球形，因此海洋的水面也呈球形，由地球引力来维持水面球形状态。另一方面，地球在自转，海洋中的水也会受到地球自转带来的离心力的作用。赤道上转速快，因此所受的离心力更大。除此之外，太阳和月亮对海水也有引力。由于月亮和太阳在不同时刻相对位置不一

样，因此海洋中每一处水所受的太阳和月亮的引力也随时间变化。正是这种变化，会引起水位高度有变化。这种由地球自转离心力、太阳和月球引力引起的水位变化，会产生重力波，导致海岸附近海水涨落，称为海洋潮汐。

由于地球差不多转半圈，太阳和月亮引起的引力变化才完成一次交替，因此海洋潮汐的波长最长可接近地球周长的一半。海洋潮汐的波长是海洋水波波长中最长的，往往比海水深度大许多倍，从而海洋潮汐是浅水波。因此，有潮汐时，我们看不到风浪那样的波纹，因为波长太长了，只能看到海水涨落。

有两个因素决定潮汐海水涨落的周期。首先，月亮和太阳的相对位置每昼夜更替一次，因此一天可以发生两次左右的涨落。另一方面，农历初一后的几天，中午太阳和月亮几乎同时处在当地天空上方（夏天甚至二者接近处在正上方），给海水的引力大，容易涨潮，到了夜晚则相反，这样昼夜引力差就较大，容易形成较强的潮汐。农历十五后的几天，地球则处在太阳和月亮的中间某位置上，到了中午和午夜，太阳和月亮对海水的引力作用方向相反，就彼此抵消一部分。但这些力的作用较为复杂，与地球、月亮和太阳在不同时刻的相对位置有关。

因此，一昼夜有两次左右的涨落，这是因为一昼夜内，太阳与月亮交替出现在我们能看到的天空与看不到的背面，与潮汐涨落当然同步。另一方面，农历初一后或十五后的几天，太阳、月亮与地球更容易在一条直线上，引起引力变化的强度大，因此这段时间潮汐强度更严重。

对潮汐这种规律的理解有助于规划出海航行。退潮时海岸面积就更大了，涨潮时海岸附近水就深了。据说，施琅收复台湾，战舰就是等到涨潮时得以靠岸。

钱塘江大潮（图2.41）

钱塘江在杭州湾出口是喇叭口形状。大海涨潮后，海水沿喇叭口向内江倒灌增高水位，形成冲击波（潮头），即钱塘江大潮。

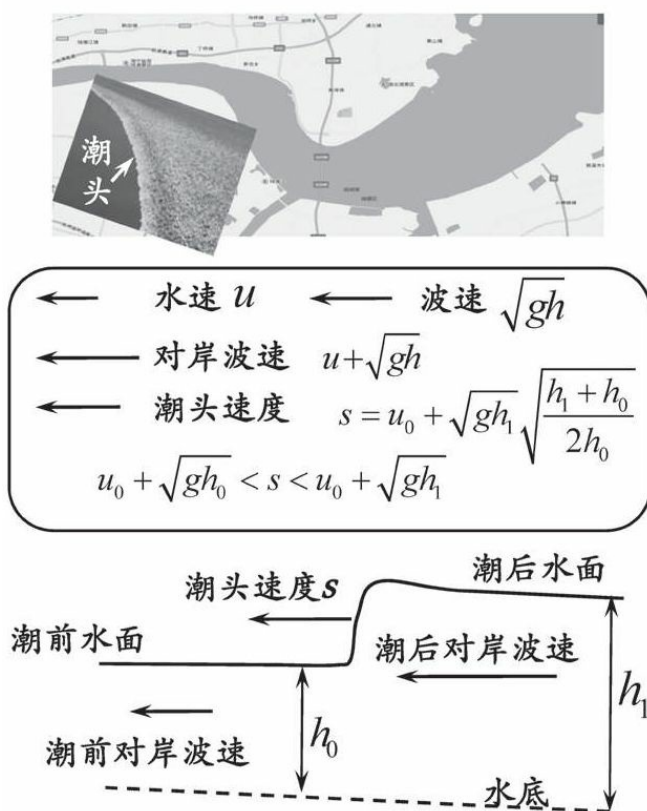


图2.41 钱塘江大潮是水位差形成的冲击波

农历八月十五后几天的钱塘江大潮，是一个极其特殊的现象。大潮的浪头并不呈波纹形状，而是一种水位高低差（数学上称为间断，物理上可以称为冲击波），潮头下游是高水位，上游是低水位。海洋潮汐看涨时，在杭州湾产生的高水位，通过重力波等方式传入较浅的钱塘江。由于钱塘江水位更浅，因此还是属于浅水波。

水位越高时，浅水波波速越大。叠加水流的对岸波速更大。高水位

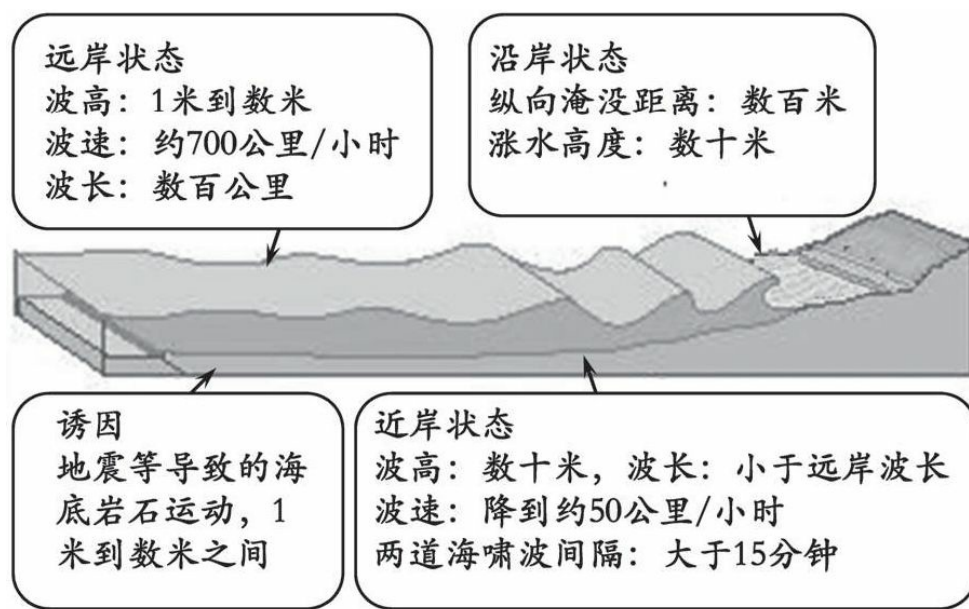
潮汐传入水较浅且呈喇叭收缩的杭州湾时，这种水位差更明显。于是，下游的高水位由于对岸波速更高，会追赶上前面的，最后挤在一起，形成潮头。

潮前潮后的水位有明显的高度差，从而两侧的对岸波速不一样。由于下游的对岸波速高于潮前的对岸波速，结果潮头的对岸速度取两者之间的折中值。即大潮浪头的移动速度在潮前潮后的两个对岸波速之间。

除了物理原因，钱塘江大潮的形成还有外因。大潮期间，月亮或太阳离地球表面直线距离近，海洋潮汐较强。当然还有其他原因。东南风也会导致重力波的产生，加强海洋潮汐。钱塘江出海口呈喇叭形，海洋潮汐的高水位被重力波等传到钱塘江后，由于越往内江水道越窄，因此潮汐水位与浪头增高。大潮在特定的江岸反射出的潮汐与主潮汐可形成交叉潮。农历八月十八，是钱塘江观潮的最佳时刻。钱塘江大潮浪高可达数米，移动速度达每小时二十多千米。

海啸（图2.42）





来源: 艾德·萨尔蒙 《海洋科学》讲稿

图2.42 海啸的形成

提到海啸, 我们会想起印尼苏门答腊岛海啸事件。海床断裂、海底火山喷发、山崩、冰川脱落成冰山等在海洋诱发的重力波, 可形成海啸。海啸对应的水波波长可达200千米。对于平均水深在3~4千米之间的印度洋和太平洋, 显然属于浅水波。以水深4千米为例, 按浅水波, 波速的平方等于重力加速度与水深的乘积, 于是可以算出海啸的波速可达到每秒200米左右, 即每小时700多千米, 比最快的船都快, 高速鱼雷也比这慢一半以上。

海啸如果在深海区形成, 那么向海岸传播时, 越靠近海岸, 水越浅, 传播速度越小, 因此后面波能追上前面的, 形成水位差(即浪头)。在海洋中央刚形成的地方, 浪高或水波波峰高度不到1米, 很难及时发现。但向海岸传播过程中, 由于海啸前的水越来越浅, 因此水位差(浪高)不断增加, 超过数米甚至数十米。这种海啸带来的水位差, 会漫过海滩, 淹没沿岸, 带来巨大的破坏。

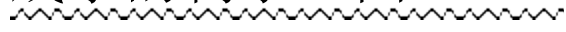
海啸在海洋中央的时候，是无害的，因为浪不高。只有到了海岸才会带来破坏。如果在岸边遇到什么海啸，那就别想到在海啸上做什么冲浪运动，只能逃走，因为破坏力太大了。为了逃离这种极具破坏力的水波，需要有预警机制，即海啸还没到达，就知道要来了，而且来得多快。

如果是深海地震诱导的海啸，那么由于地震波传播速度达每秒几千米，即比海啸快许多倍，因此会先探测到地震，据此可对海啸进行预警，及时疏散沿岸人群，减少人员伤亡。

4. 物体与水波竞速 O形波与V字形波

小物体、小昆虫和小鱼戏水，会产生涟漪，涟漪圆环形波纹的半径越来越大。小物体、小昆虫和小鱼也可能同时在朝水面的一个方向移动。如果移动速度比涟漪波速慢，那么就被套在先前激发的涟漪内。如果速度比涟漪波速快，那么就会冲出先前激发的涟漪。现在就来看看各种情况下，不同时刻发出的涟漪所排列的形态。它们能分别排出O形波与V字形波。

戏水的鸭子（图2.43）



设想一只水鸭在水面极其缓慢地移动。水鸭激起的涟漪，一圈圈波纹总是在水鸭外面，即形成套环形态。

如果水鸭稍微游快点，那么从水鸭右侧有一道向右后方直线延伸的水浪，从水鸭左侧也有一道向左后方直线延伸的水浪。两道水浪的前锋连线，形成以水鸭头部附近某点为顶点的V字形。

行走中水龟的涟漪 (图2.44)



图2.43 水鸭激发的水波

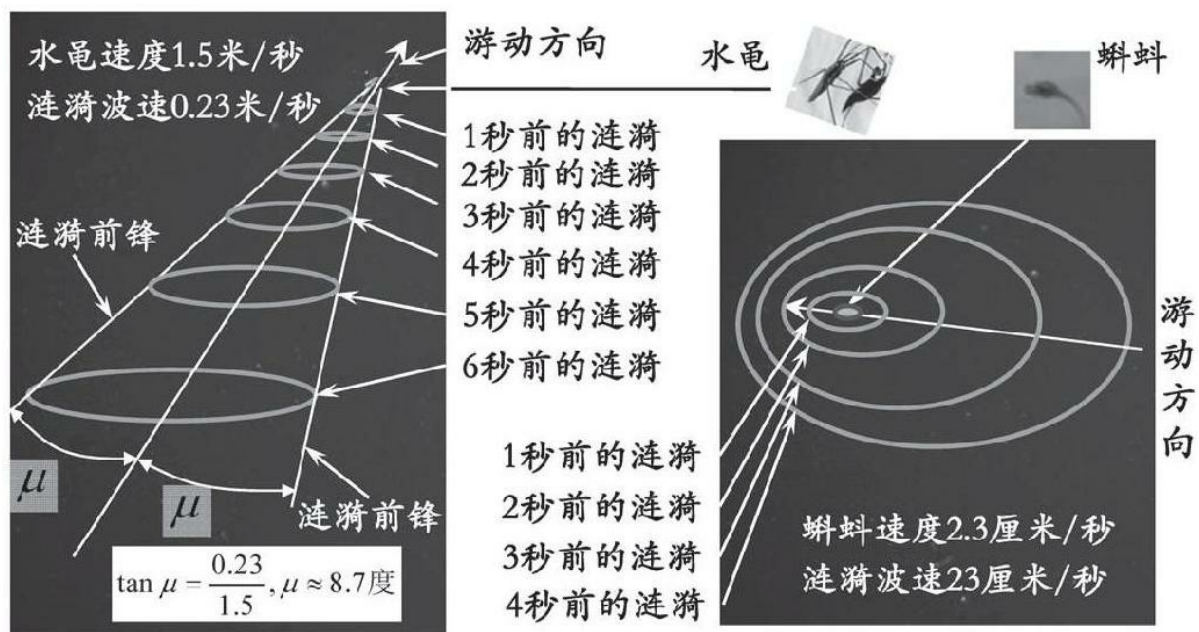


图2.44 较快的水龟与极慢的蝌蚪引起的涟漪圈圈排列方式

水龟最快滑行速度为每秒1.5米，即百倍于身体长度的距离。水龟激发的涟漪圈圈传播的速度比每秒23厘米快不了多少。假定水龟以每秒1.5米的速度走直线，且不断激发新的涟漪圈圈。

水龟速度比每个涟漪圈扩张的速度快多了。于是，前期时刻产生的涟漪圈圈，都落在水龟后方，圈圈无法跑到水龟前方。各时刻发出的涟漪圈圈，正好夹在一个V字形的中间。水龟跑得越快，V字形的顶角就越小。水龟只对这个V字形内的水有扰动，V字形外的水没有受影响。

这种扰动范围被局限在V字形围成的区域，与超声速飞行引起的声波的影响区域完全类似。超声速飞行中这种V字形波是声波前锋的边界，称为马赫波，V字形的半顶角称为马赫角。如果在小池塘看水龟行走，就会看到涟漪圈圈的这种排列方式。

打水漂时，水漂在各时刻弹跳入水激发的涟漪，也像水龟的涟漪一样，前锋连线也是V字形。

蝌蚪的涟漪（图2.44）

蝌蚪比水龟慢多了。假定一只小蝌蚪以每秒2.3厘米的速度游直线，每隔一秒冒出水面，产生涟漪。由于蝌蚪游动的速度比涟漪外圈至少是每秒23厘米的扩张速度慢，因此早期时刻的涟漪圈圈总是包含晚些时候的涟漪圈圈，蝌蚪总是在所有圈圈的里面。这就是套环形态。

亚声速飞机激发的声波，也是这种套环形态。

奇怪的水鸭V字形波（图2.43）

游动速度高于波速，就会把圆环形波浪限制在V字形围成的区域内，这种情况称为超波速运动，如同飞机超声速飞行。如果低于波速，按理波纹形成套环形态，这种情况称为亚波速运动，如同飞机的亚声速飞行。

水鸭形成的V字形波并没有像水龟或水漂那样夹带一个个圆环形波纹，而是沿着V字形的边缘，出现两排长度较小的短波浪。每排小波浪，好像是V字队形的候鸟在迁徙一样。它们是如何形成的，为何乖乖地在那里排队？

原来，对于水龟和蝌蚪这样的小生物，激发的涟漪，对水面只是一种小扰动，波纹形态完全由不同时刻的涟漪简单排列在那里，相互之间没有干扰，就像在纸上画圈似的，也不会引起其他水位的变化。可是，厚实的鸭子就不一样了。鸭子游动较快时，会把前面的水位推高，推出碎浪。这如同大风吹出的杂乱的风浪。

风浪中长出有形的涌浪，分离出来，往一个方向传播。类似的，鸭子激起的波浪，也长出小涌浪，它们有自己的速度，形成了一节节小波浪。奇怪的是，不管鸭子速度如何，V字形波的半顶角都是一个给定的值。过去认为，这个值与游船激发的V字形波的半顶角是一样的，因为鸭子也像一条船。具体情况，我们还是在介绍船波时再讨论。

5. 破浪而行 船与游泳形成的波浪

船在乘风破浪航行时，显然会把前面的水位推高。这种被推高的船头波会下蛋似的，吐出一道道较短的小波浪。小波浪们从船的两侧，排

成V字形，如同候鸟迁徙一样。游泳运动员由于速度较快，也会在头部推高水位。挡住了前方的空气怎么呼吸啊，别担心，侧脸就是低水位，那里可以吸气。100多年前，开尔文发现，船边形成的V字形波的半顶角都是 19.47° ，不管船的速度如何。因此人们把这种波称为开尔文波。可是，开尔文波难道与超声速飞行中声波的前锋面组成的波（称为马赫波）没有关系？如果有，后者的半顶角可是与飞行速度有关的。

开尔文波 排队的小波浪（图2.45、图2.46）



图2.45 V字形波浪与候鸟V字形迁徙

与水龟和蝌蚪不一样的是，船沉入水中的深度较大，因此船体撞击挡路的水，会推高水压。一方面迫使部分水从侧边流过，另一方面迫使水位提高。这种抬高的水位与前方的水位差，就是船浪，也称为船头波、冲击波和艏波。

除了船头波，还有船边白浪。白浪一边翻腾一边破碎。然而，乱中有形，船边碎浪会从两侧吐出一条条小波浪，规矩地排成V字形。

一百多年前，开尔文就发现，这个V字形波的半顶角是 19.47° ，与船的速度无关。

当然，也不单是船才会产生船头波。游泳选手游泳时，在头部也会产生冲击波。波浪的波峰离开头部有一点点距离，但你不能对着这个波峰呼吸，否则一浪袭来，就会呛水。波峰的下游尤其在头部的侧面有一个低水位的波谷，这可以帮助游泳选手呼吸空气。如果没有这种波谷，就很容易将水吸入呼吸道。当然，将头调到侧面，波谷陷下去的空间更大，这样更有利于呼吸，而且不会直接撞到水。



图2.46 船头波、船尾波与游泳波

船边碎浪看上去没有什么形状，但会在两侧产生一段段圆弧形小波浪。一个挨着一个，像排队一样。两侧的队列正好构成一个V字形。小波浪们的这种排队方式，如同候鸟迁徙一样。候鸟之所以排成V字形队列迁徙，是因为可以省力。

从船头边碎浪分离出有形的小波浪，道理如同风浪中分离出规则的涌浪。其实，碎浪中包含了各种各样的波长的波，有的跑得快，有的跑得慢。那些跑到最边上的小波浪们，是那些能从碎浪中吸收能量的波。

至于为何小波浪沿着半顶角正好为 19.47° V字形边缘排列，找不到一种很直观的方式来理解。开尔文就那么伟大，用理论证明了这个结果。

用波长除以周期，得到了一个波速，这个波速称为相速度，就是波形（专业上称为相位）的移动速度，是形态的移动速度。

小波浪却不是来源于相速度的传播。可以用给救护车让道的例子来理解，虽然不是特别严谨。在弯弯曲曲的山道上，汽车一辆接着一辆行驶。从高空看汽车沿着山道走波浪曲线。设想一辆救护车从前方过来。每一辆车见到救护车靠近，都会往路边方向移动一点距离让道。于是，从高空看，车辆连成的曲线在救护车行驶的位置，有一个小的弯曲。这个小弯弯形态的移动速度，完全由救护车的位置决定。这个由救护车决定的车流形态的移动速度，称为群速度。上面提到的小波浪们，就是以类似的群速度移动的结果。

开尔文角被马赫角取代 从水面走向空中（图2.47、图2.48）

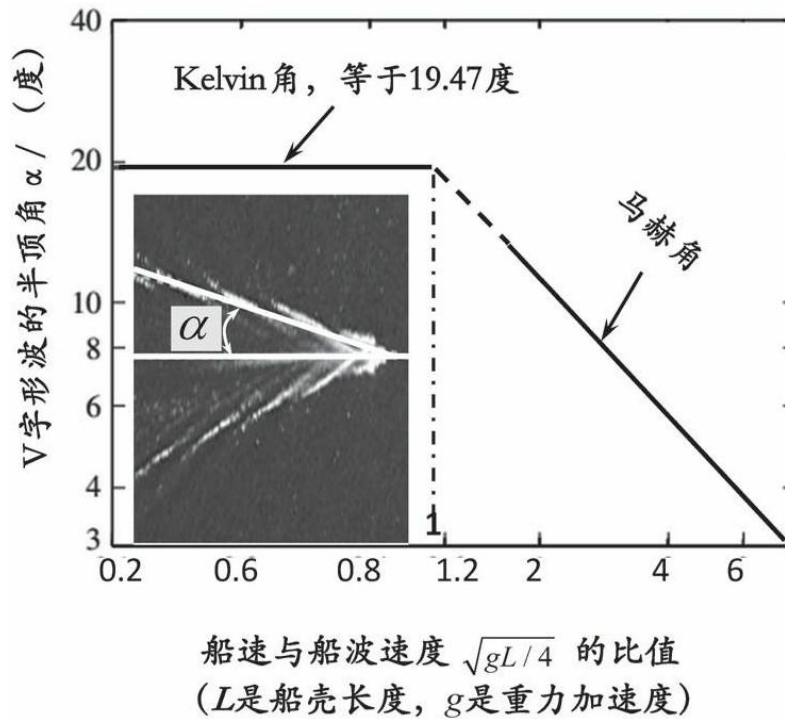


图2.47 V字形波的半顶角：开尔文角与马赫角

巧合的是，开尔文角与超声速飞机的飞行马赫数恰好为3时，形成的马赫波的马赫角一样大。这个并不表明船的航行马赫数就是3，机制完全不一样。

可是，这一结果在2013年有了新的发现。法国科学家拉波德与莫瓦斯得出了这样的结论：对于一艘给定长度的船（或长方形物体），如果船速小于船引起的重力波的波速，那么V字形波的半顶角就是开尔文角，不管船速如何都是这个角。如果船速高于该波速，那么V字形波的半顶角就是马赫角，即与船速成反比。于是，开尔文角（19.47度）就是半顶角的最大值。

船速高于重力波的波速时，如同超声速飞机的飞行速度高于声波的速度。因此，这一发现将船波与超声速飞行的马赫波统一起来。船边的小波浪排成V字形，与候鸟迁徙排成的V字形队列也相似。这些相似

性，有的来源于道理相通甚至相同，有的就是一种巧合。

水中世界还有许多，比如说各种各样的水下生物的运动。在这里不能过多留恋水下世界，因为空中的秘密同样精彩。

那就让我们像蝠鲼（也称为魔鬼鱼）一样，从水下跳到空中吧。我们纳闷的是，蝠鲼虽然具有完美的流线型，不亚于协和飞机那样的优美外形，给它取的名字却这个也不好听，那个也吓人。

跃到空中滑行



在水下像飞机一样航行

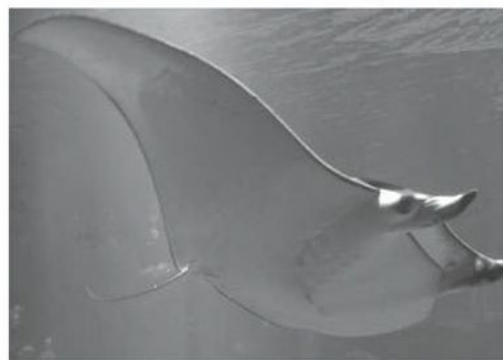


图2.48 像蝙蝠侠一样的蝠鲼

第三篇

空中旋律

空气是看不见的气体，空气的分子在随机躁动，这些躁动给我们带来温度，也带来气压，也能让其中的氧气与化合物结合燃烧释放能量。物体与生物在空气中运动，空气会和谐地膨胀与压缩，让翅膀拥有升力，让落体遇到阻力，让声音得以传播，让发动机拥有推力。气压随时间的快速变化带来声音，和谐地变化带来音乐。声波的传播像田径接力棒运动。气压在物体上随位置的变化带来升力，使飞机能在空中远航、鸟能在空中飞翔、昆虫能在空中悬停、旋转的叶片能产生推力。

海洋蒸发的热蒸汽与空气携手，在地球自转的协助下，产生台风那样的自然奇观。风雨雷电与云彩，让天空美不胜收；风吹雨打，让生物茁壮成长。大风让田野出现麦浪，让大海波浪滔滔。火旋风举着长长的火焰，会把火种传得很远。空气能与自由的小物体共舞。随手释放一个名片，空气使它会翻滚着下落。枫叶种子螺旋般地掉落时，会搅动起旋涡。

空气很轻，速度让飞机能飞起来。空气不重，浮力让气球能飘在空中。空气挡住了宇宙射线，通过把自己牺牲成臭氧减弱有害的紫外线。它义无反顾地烧尽大部分坠落的危险太空碎片。

3.1 绚丽多彩与刚柔并济的地球大气

抬头看天空，除了云彩和星星以及偶尔出现的流星与幽闪，似乎什么都没有。其实，地球周围充满了空气，我们称之为地球大气。大气有时也安静，地球引力不让它们飘往太空，阳光照射让它们拥有温度与气压，海洋蒸发让它们拥有水分。温度、气压和水分既让大气有生气，也让大气有情绪：风雨雷电，冷暖无常，这种情绪状态与变化也称为气象。万有引力、阳光的照射、地球的自转、海洋的蒸发，让大气充满变化。气压的变化，可引起让地面建筑物摇晃的大风。大气密度的差异，让不同速度或类型的飞行器分享不同的高度。温度的分层效应，让大气可以上下对流，也可以让污染物被锅盖一样的逆温层罩住。分子高速热运动的接力棒效应居然让声波每秒能跑数百米，惊雷不仅响彻云霄，而且比声音跑得快。声音如果每秒只能传播几毫米，听你一句话得等十分钟。轻轻的空气比水还容易旋转，尘卷、火旋风、龙卷风、台风，大气并不柔情似水，但至少有点暴力美。

1. 大气的力量

大气中的空气分子很轻，但一起发威时力量不可小觑。它掀起的台风连树枝都可以折断。当然，大气的运动也可以让天空彩云朵朵，让红旗飘飘。大气与水相比，虽然那么稀薄，但足以让昆虫、飞鸟和人造飞机能在空中飞翔。你实在觉得空气在给你提供升力的同时也会施加阻力，那就跃到大气层之上，像卫星那样，用轨道离心力平衡地球引力。不过，请注意，那里空气太稀薄，因此不要把头伸出去。

风的力量（图3.1、图3.2）

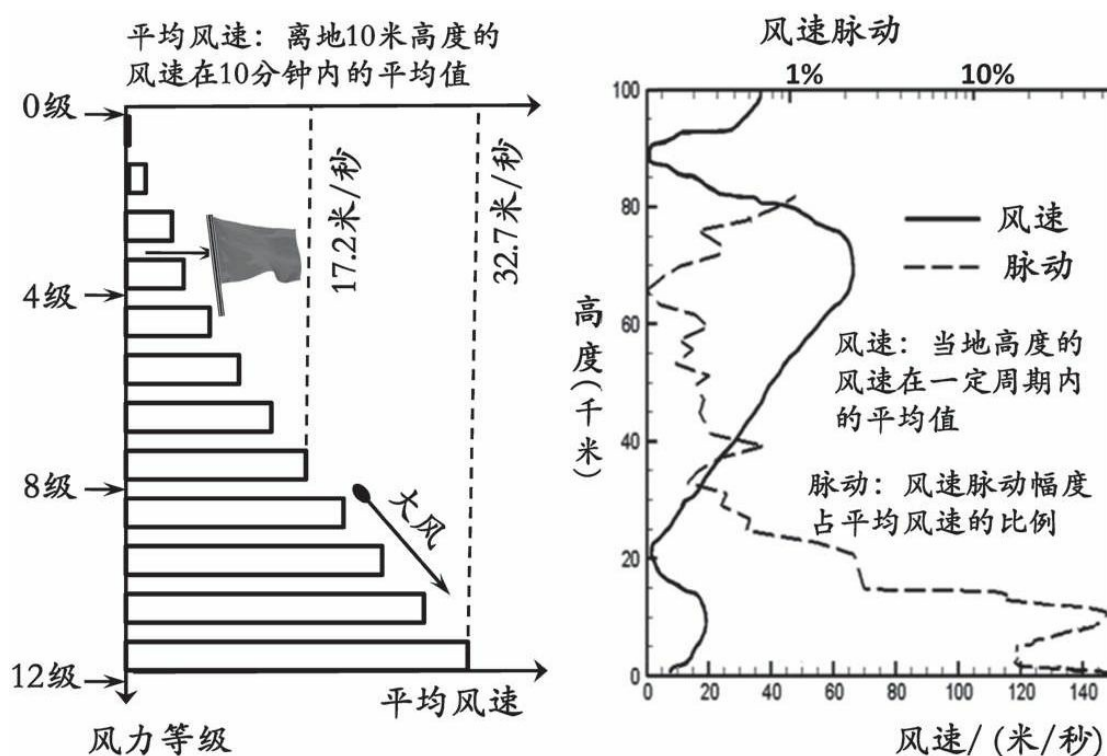


图3.1 风力等级变化与风力随高度变化

大气中有风雨雷电，这里我们只看风的力量。风来自于不同地区的气压差。这些气压差可能源于温差或者地球自转等因素。温差来源较多，大气随高度变化有温差，赤道与北极之间有温差，海洋与陆地之间有温差，地貌的高度差异也带来温差。

气压差以声波传播，但一处的风传播到另外一处的速度可能小于声波的速度。不同强度的气压差导致不同的风速。地面由于摩擦，风速低于高空风速。离地10千米左右，平均风速有一个极大值，接着风速随高度下降，到了20千米左右，又有一个极小值，极小值的平均风速可能只有每秒几米的量级。70千米左右又出现一个极大值，90千米左右出现一个极小值。大气的风速随时间也有变化，实际上有大气湍流脉动，即围

绕平均值，风速大小杂乱无章地变化，变化幅度与平均风速相比，在离地10千米高度以下，可以达到10%的量级。

陆地上有炊烟、树木花草、建筑物，还有动物。零级风时，炊烟在浮力作用下，像立柱一样向上升起。

不同速度的风，会让这些不同大小的物体受的风阻不一样，从而运动状态不一样。从它们的运动状态，也可以判断风力的大小。风可以把尘土吹起，把秋叶吹落，让蒲公英的种子远距离播撒，让建筑物摇晃。设想我们的人体迎风面积为1平方米，风阻系数是1，那么八级大风就可以产生20千克左右的风阻。如果不站稳，就很容易被吹倒。

风阻来自于气流作用在物体上的摩擦力和气压差。风阻对物体运动的作用与风力大小密不可分。

一级风时，炊烟随风向摇摆，即烟的摇摆方向能指示风向，树叶和风向标却静止不动。二级风时，外露的皮肤会有风拂面的感觉，树叶能发出轻微的响声，旗子开始飘动，陆地上的风向标会随风移动，能指示风向。三级风时，树叶和很细的树枝摇动不息，且旗帜会展开。

到了四级风，灰尘和软纸张随风飘起，小树枝在摇动。五级风时，小树枝会移动，带叶小树会摇摆。六级风情况下，大树枝在摇动，电线呼呼有声，撑伞非常困难，地面上的塑料箱会被吹倒。到了七级风，整棵树会摇动，迎风步行很费劲。



图3.2 风中的波浪云与红旗飘飘

在八级大风中，一些小树枝折断，迎风步行感到阻力很大，汽车在路上被吹偏，行人步履维艰。九级风能将一些树枝从大树断开，整棵小树被从地拔起，建筑物上的临时广告牌和栏栅被吹走。十级风能让树木被吹断或连根拔起，建筑物可能损坏。陆地上大量的植物和建筑物会遭十一级风损坏。

到了十二级台风，陆地上大量植物和建筑物遭损坏，碎片和未加保护的物体被掷起抛向空中。

有时可能出现一股不同风速的风从高空平行地吹过。我们从失稳的多姿多彩中已经知道，如果大气中有两股风速不一样的风平行流动，那么就相当于速度更快的风在吹另外一股风的空气，失稳导致出现开尔文波浪云。这种不稳定使旋涡卷起，加速了两股风的掺混，更快地缩小了两股风在交界面附近的差距。也可以说，因为缩小差距是演化的必然趋势，因此必然出现大气波浪。

大气波浪的旋涡通过离心力等因素消耗涡心的气压，使潮湿空气凝结成美丽的波浪形云朵。

旌旗在风中，两面都会受到风的吹拂，也会产生波浪，旌旗舞动就是两面都用风吹的结果。这种波浪的形成，和大风之下水波的产生原因很类似。

地球大气与飞行（图3.3、图3.4）

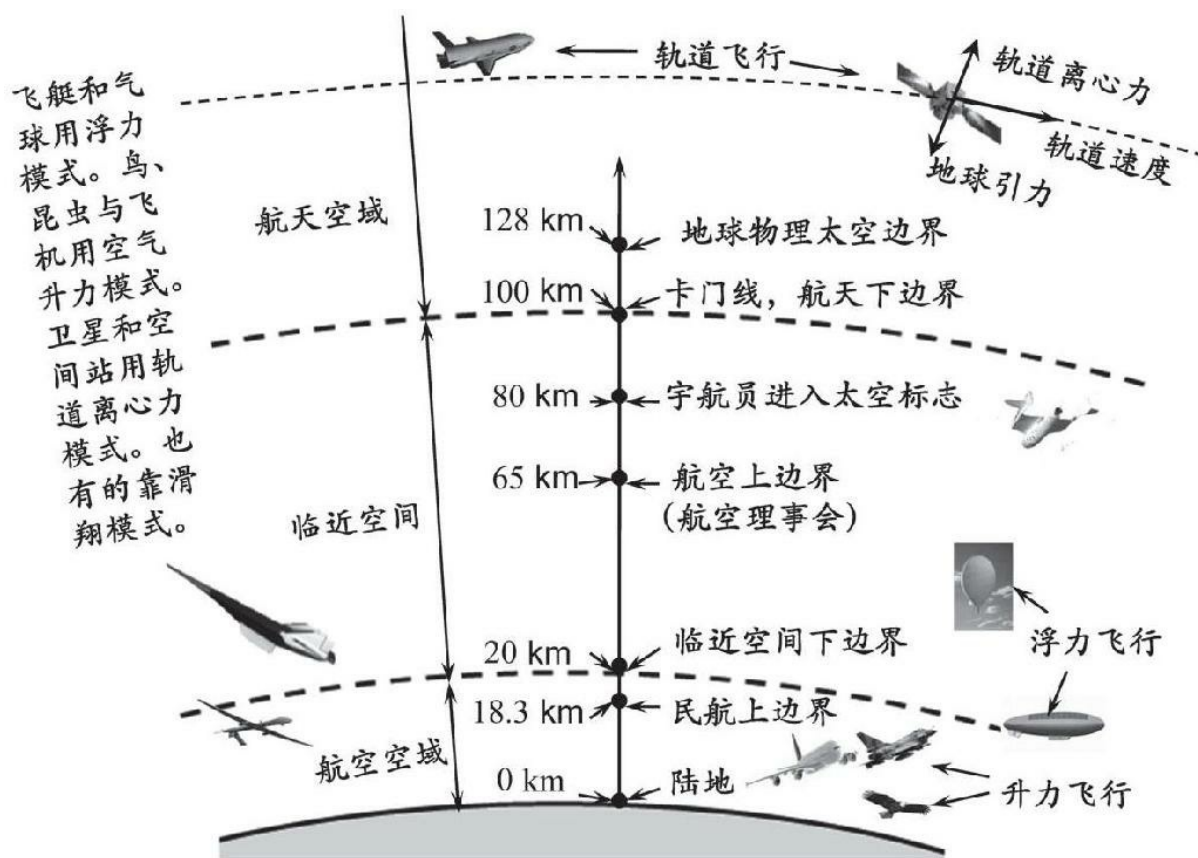


图3.3 飞行空间与飞行模式

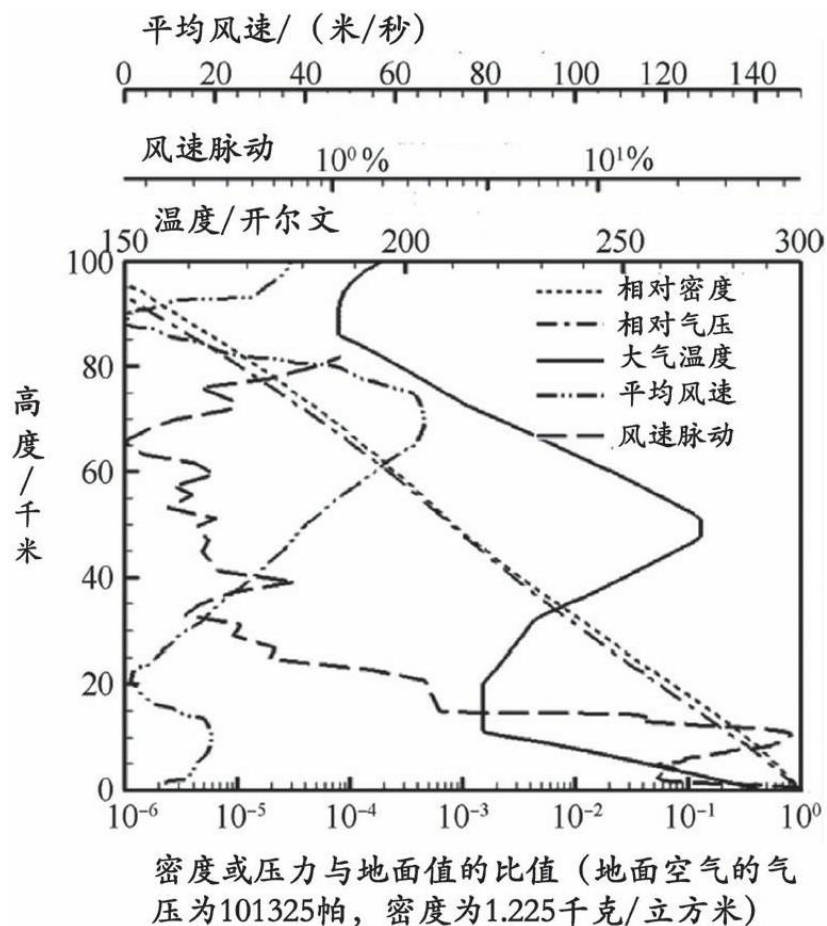


图3.4 大气参数的平均值随高度变化

我们在地面，只需要关心温度和风随季节的变化。但对于飞行专家，就需要关心大气参数随高度的变化。这些大气参数决定了什么样的飞行器怎样飞行，能飞多高、飞多快、飞多久。

正是地球引力，将地球周围的空气分子捆在大气层中。地球大气中的大气分子，靠分子热运动产生内压，抵抗地球引力。正因为如此，越靠近地面，大气压力越高。地面空气的密度是每立方米1.225千克，气压为101325帕（每平方米1千克的力就是9.8帕，这个9.8就是靠近地面的重力加速度）。到了20千米的高度，空气的密度只有每立方米0.088千克，气压为5474.9帕。到了32千米的高度，空气密度为每立方米0.0132

千克，气压为868.02帕。

因此，大气密度与气压随高度下降。离地越近，气压越大，这样才能支撑起上方的空气。或者说，正是这种高低气压差，平衡了两个高度之间的空气的重量。气球之类的物体放在大气中，气压差就提供了浮力，因此气球与飞艇之类的，是靠浮力飘浮。浮力恰好等于排开的空气的重量，因此携带比空气轻的氦气或氢气的气球，能浮在空中。

依靠机翼产生升力的飞机，其升力大小既正比于机翼面积，也正比于飞行速度的平方和大气密度。如果飞很高，大气密度的下降会降低升力。此时，要么增大飞机机翼面积，要么增大飞行速度。

基于以上原因，不难理解什么高度适合什么样的飞行。

大多数昆虫由于翅膀太小，只能在离地数十米以下飞行，以免被风刮走，并且能在较稠密的地面附近空气中能产生足够的升力。大多数鸟也只能在离地数百米以下的高度飞行。据说翅膀足够大的老鹰可以飞到10千米左右的高度。

直升机一般不会超过六千米的高度，因为是靠旋翼产生升力，转得太慢升力不够，转得太快容易损坏并且翼尖容易超过声速（超过声速后容易产生不利于升力且会增大阻抗的冲击波）。

一些军用飞机可以飞到10千米以上。民航客机一般在10千米左右以及以下的高度飞行。10千米左右高度的大气湍流较为强烈。飞机遇到强烈的湍流，就会颠簸，这个高度的平均风速出现一个峰值。但大部分民航机都在这个高度飞行。因此，我们在飞机电视屏上会看到对空速度和对地速度。对空速度是飞机相对于静止空气的速度，对空速度是多大与飞机的飞行能力有关，因此是设计速度。对地速度是飞机速度减去或加

上了当地的风速。虽然如此，飞机的飞行速度在好几百米每秒，且飞得越高速度越快，因此风速与飞行器速度相比非常小。

20千米左右高度的平均风速有一个极小值，因此20千米左右有个低速风带。人们希望研制出高高空飞艇，驻留在20千米左右的高度。在那个高度，因为风速低，不容易被吹到别的地方去。如果维持在那里不动，只需要小小的动力就可以做到。可是，那里空气密度太低，需要巨大尺寸、长度100米量级的飞艇才能拥有足够的浮力。这么巨大的东西，发射、回收和控制都很难。一旦上去，就不要指望马上下来，而长久驻留不能靠烧油，只能靠太阳能电池接收太阳能，将其转换为电能，电能驱动螺旋桨提供推进。

离地18.3千米是国际民用航空组织规定的民航上边界。据说，未经许可，飞到别的国家18.3千米高度以下的空域，称为侵犯了领空。国际航空理事会规定的航空上边界是65千米。如果能飞到80千米，就算接近了太空。因此，离地80千米被作为宇航员进入太空的标志高度。虽然不同行业的规定有一些差异，但目前一般把20千米以下的空域称为传统航空空域。

在20千米左右高度的空气密度已经很低，只有地面的十四分之一左右，因此为了获得足够的升力，无人机的翅膀要做得很大。在30千米左右的高度，空气密度只有地面的百分之几，因此这里适合高超声速飞行，通过马赫数5以上的超高速，获得足够升力。如果到了100千米的高度，密度是地面的百万分之一量级，这么稀薄的情况下，只有飞行速度达到每秒7.9千米以上，才能产生足够的升力。而这个速度就是做轨道飞行的最低速度。

所谓轨道飞行，就是卫星那样的航天器，绕地球做圆周运动。由于

走弧线，就产生了离心力，可以平衡地球引力。如果用轨道速度飞行，那么离心力就可以正好平衡地球引力。既然离心力可以平衡地球引力，那么就不需要机翼，不需要升力了。因此，100千米以上的高度，适合轨道飞行。因此，把100千米以上的空域称为航天空域，或者100千米称为航天的下边界。由于这是卡门计算得到的结论，因此也把100千米的高度线称为卡门线（应该称为卡门球面）。航天空域适合航天器做轨道飞行。

要飞到20~100千米之间的任意高度，像无人机那样靠增大机翼面积，高超声速飞机靠增大飞行速度。由于密度和气压低，普通发动机很难直接从大气中提取能达到燃烧气压与密度条件的氧气，很难产生足够的推力，除非研制基于新的原理的发动机。由于这个高度的特殊性，因此人们把20~100千米这个高度范围定义为临近空间。这个空间适合发展长航时无人机和高高空飞艇这些可以利用太阳能的飞行器，也适合发展高超声速飞机。

2. 神奇的自洁功能 环境污染

地球大气两头受虐，太阳光线的紫外线从顶层侵入，人造污染物从地表升起。大气中的氧分子横刀立马，通过将自己的一部分牺牲成臭氧，吞下了大部分凶悍的紫外线，并让海拔50千米以下的一层内，上热下冷从而气流平稳。这一层以下的大气继续吸收阳光的其他部分，将无害的一部分留给地球表面。地表知恩必报，把一部分转变成热量，馈赠给近地大气，加热了地表大气。这如同锅底烧水，会让地表附近的大气沸腾。也同时，沸腾带来的大气上下对流，让风雨雷电有了舞台，也能将人造污染物被吸往更高的空间得以稀释。可是，当早晚没有阳光时，或当高山送来暖风时，地表附近本来上冷下热的温差就可能被逆转，于

是沸腾停止了，或者盖上锅盖了，污染物此时会借机在低空施虐。库兹尼茨环境污染曲线告诉我们，我们彻底战胜污染，也得有足够的经济实力。

温度分层与大气沸腾（图3.5）

与大气压力和密度随高度上升单调下降的规律不同，大气平均温度随高度的变化不那么单一。原来，大气吸收阳光，在不同的高度发生不同的吸收过程，决定了温度分层分布。

太阳中有紫外线，波长越短对人体越有害。万物就是那么奇特：越有害的紫外线，因为波长越短，频率越高，所以能量越高（光子的能量与频率成正比，比例系数称为普朗克常数），与大气中的氧分子碰撞时就越容易自毁。这让我们在地球上能免受紫外线的危害。

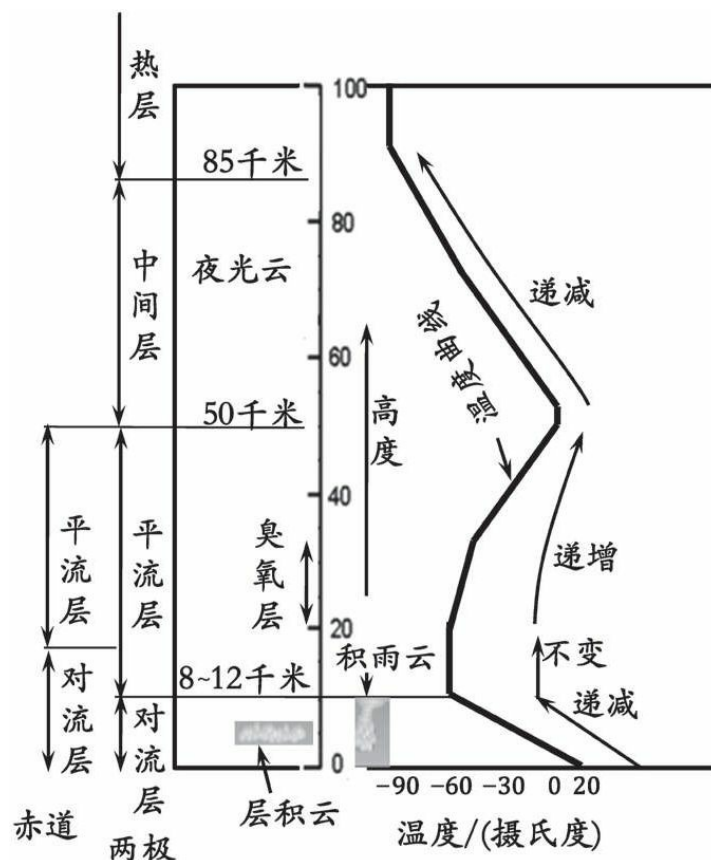


图3.5 大气分层与温度随高度变化

离地50千米以下的氧气已经足够多，遇到紫外线，它们甘愿粉身碎骨，一部分义无反顾地分裂成氧原子，再与氧分子结合成可进一步吞并残余紫外线的臭氧（三个氧原子组成的分子是臭氧）。吸收了紫外线的臭氧会还原成氧分子和氧原子，同时把吸收的紫外线的能量变成热能，加热平流层，让平流层下冷上热从而气流平稳。这样形成的抵御紫外线的臭氧，在20千米和30千米之间浓度达到最大值，因此那里也称为臭氧层。氧气这种变有害为无害的能力，让我们感叹大自然为何具备这种巧夺天工的神奇力量。被摧残的氧分子身躯断裂，一半依旧依附在其他氧分子身上，继续战斗，像神盾一样挡住了来犯的紫外线，让地球人类免受其害，却还落个臭氧的污名。你不能仅仅因为不习惯它的味道，就在前面加个臭字。哪怕最不尊敬，也可以称其为三原子氧吧。它们可是为

了保护你而变成了那样。

太阳光到达大气层，约29%被反射回太空（云、大气分子和地球表面的冰与雪等参与反射），约23%左右被大气吸收（水汽、尘埃、臭氧和大气分子等能吸收阳光），48%穿过大气到达地球表面，被地球表面吸收，从而加热地球表面。物质吸收阳光能量后，其中的分子热运动就变得更活跃了，于是就有了温度的提高。温度提高到一定程度后，又会将一部分能量辐射出去。这种辐射满足这样一个规律：物质温度每增加一倍，其辐射出去的能量就增加到16倍。正因为这种辐射，吸收太阳光的地球与大气的温度才不会无限制地提高。

因此，被加热的大气和地球表面均向外辐射能量。大气能辐射掉吸收的阳光的59%，而地球表面只能辐射掉12%（大气是全立体空间位置的辐射，地球则只从表面辐射，当然辐射掉的少）。因此，地球表面温度比大气温度高，从而会加热靠近地面的大气。这种地表加热效应，使越靠近地面，（平均）温度越高。

于是，在靠近地表的一层大气内，上冷下热（可以称为顺温分布，虽然人们并不这么称呼）。我们用火从锅底加热烧水，一定时间后，水就沸腾了。沸腾的水中，有上下对流，快速将锅底高温散发到水表。空气很轻，比水更容易沸腾。因此，在顺温的作用下，也会沸腾。这就引起上下对流，像沸腾的水一样。

在大气沸腾的作用下，靠近地面的一层是对流层，即空气有上下对流。你可以将一根很轻的小毛毛扔到空中，偶尔能看到毛毛直蹿到高空，就是因为有大气垂直对流。

对流层在地球南北极能到达8千米的高度，在地球赤道则达到18千

米的高度（因为赤道附近阳光照射更强，这种温差导致的沸腾效应就更强），平均而言，对流层从地表延伸到11千米左右的高度。

对流层主要是顺温分布，即越往上温度越低，而平流层是逆温分布，越往上温度越高。它们之间有一个过渡层，称为恒温层，温度不随高度变化。

大气污染的扩散（图3.6、图3.7）

大气分子有热运动。热运动速度指向各个方向，可以达到每秒数百米。烟气中的粒子也很小，比如说微米量级，会遭遇空气分子的碰撞，也会参与热运动，于是也跟着乱蹿。这就是扩散效应。

扩散效应使得烟气柱越来越宽

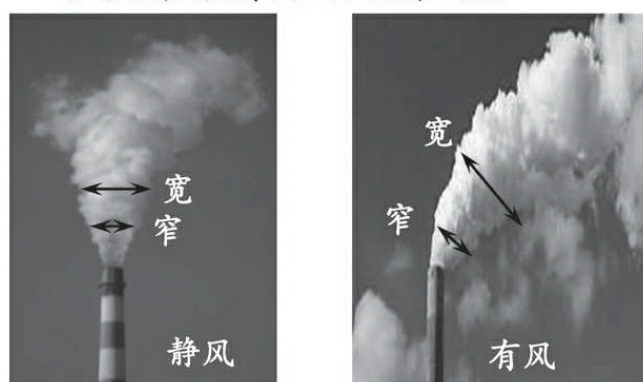
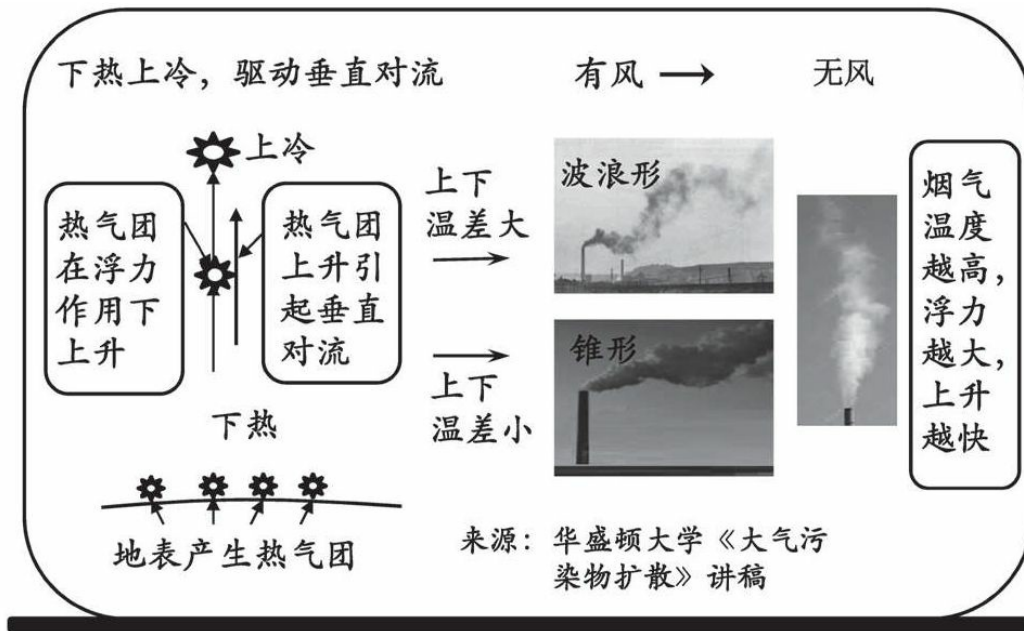


图3.6 大气扩散将污染物分布到更广的空间

这种扩散使烟柱越来越宽。如果是静风天气，那么烟柱上升的同时，其宽度越来越大。如果有风，那么烟柱会斜着上升。

大气中的湍流脉动的扩散能力比分子热运动更有效，可将扩散速度加快近许多倍，据说最高可加快近百万倍。



正常气温引起垂直对流，将地面大气污染稀释

图3.7 上冷下热时污染物的扩散

顺温（即上冷下热）引起的大气垂直对流也会携带烟气一起上升，使污染物向高空和远方加速输运和扩散。这就是对流层具有对低空污染物自洁作用的原因。

当然，烟气出口温度如果较高，那么其自身还有浮力，浮力大小与自身温度和环境温度差有关。出口温差越大，进入大气后浮力越大，上升越快。上升的高度满足布里格斯（Briggs）模型，该模型指出，上升高度正比于温差的三分之一方。

因此，提高烟气出口温度，有利于污染物向高空扩散。

逆温层效应与污染加剧（图3.8）

对流层内上冷下热，温度随高度上升而单调下降，但这只是平均而

言。早晚没有阳光照射情况下，辐射会使地表和大气的温度有所下降。地表比大气辐射能量快，因此早晚降温更快，这可能使地表温度比附近大气温度更低，从而使在某高度范围出现上热下冷的状态。这种现象称为逆温现象，即本来上冷下热的温度分布逆转成了上热下冷。

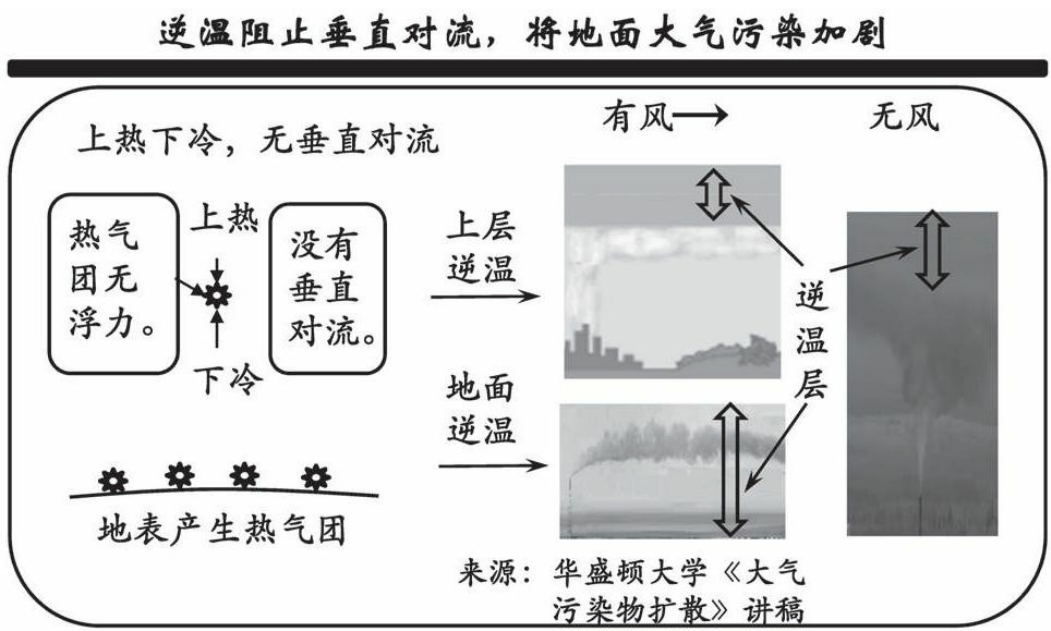


图3.8 逆温条件下的污染

烧水时，如果在上面的锅盖位置加热（比如说用电加热），那么锅盖下的水绝对不会沸腾起来。逆温层也是这样，由于上热下冷，因此逆温层内不会有沸腾，也就不会有上下对流。

有逆温层时，由于没有垂直对流，单纯的分子扩散就不容易将污染物扩散开。夜间形成的逆温层，靠近地面，从而地面污染物不容易扩散。因此，尤其要避免夜间排放。

如果周围地势高，或者有高山，那么从高处过来的暖气流，或者从高处下压过来的气流，会在头顶某高度范围制造一个薄薄的逆温层。这种悬在头顶的逆温层，如同锅盖一样，下方沸腾的空气无法蹿到上面

去。污染物也无法穿越该逆温层，故容易造成局地污染加重。

因此，应该避免或减少在低洼地带或山坡下排放污染物。

库兹涅茨环境污染演化规律（图3.9）

环境污染库兹涅茨演化曲线

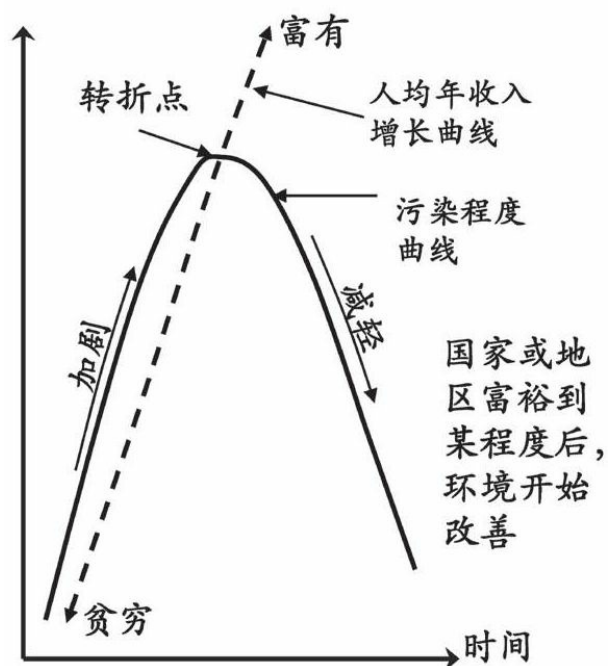


图3.9 环境污染与富裕程度

库兹涅茨环境污染演化规律指出，社会由贫穷变为初步富有的过程中，会制造越来越多的污染，导致环境污染程度加剧。

富裕到一定程度后，才开始意识到或者有条件清理环境。

按购买力平价计算，人均收入在1.7万~1.8万美元时，应该出现转折点，之后环境会逐步改善。

当然，制造了多少污染与环境可观测到的污染是有差别的，与气象条件有关。某地在正常天气时，污染物可扩散到数千米高，显得污染不重；而出现逆温天气时，只能扩散到数百米高，聚集在地表容易形成雾霾天，显得污染很严重。

假设1000千米以下是雾霾，污染指数是500（即500微克每立方米）。地表每平方米上，1000立方米的范围，这种污染物的总的质量只是0.5克（1000立方米乘以500微克 / 立方米）。于是，地面每平方米只需要0.5克的尘土（大概半粒花生米的质量），变成极其细小的粉末扩散到1000千米高度以下范围内，就可以造成污染指数爆表。可见，保持裸露的地面一尘不染多么重要。

美国经济学教授泰勒·考恩认为，由于中国人均年收入按可比价格已经达到了1.4万美元，因此距离改善环境不远了。另外，像北京这样的发达地区，2016年比2015年的环境有所改善。

将经济结构朝着低能源消耗和高节能的方向发展，提高办事效率减少不必要的出行与活动，优化地理与施工环境，这是从战略层面解决源头问题。减少排放过程中的污染物，这是从技术层面解决问题。避免夜间排放，提高烟气温度，避免在低洼低坡地带建厂排烟，这是从管理层面解决问题。如果依然有污染物飘到了天空，是否被逆温层困住，是否被形成雾霾，就只能听天由命了。

3. 分子的运动 个体的行为与集体的表现

空气的每个分子有自己的个体行为，就像每个人一样。每时每刻，每个分子古怪地在做杂乱无章的运动且相互撞击。这似乎是内斗，但这种内斗的集体表现让大气拥有适合的温度与气压。它们在一点的随时间

变化会转变成波浪（压力波），被相邻分子用接力棒方式快速传递从而引起周围的空气共舞，就像墨西哥人浪似的或者水中涌浪似的。正因为如此，声音得以传播。气压在空间不同区域的差异，又会迅速转换为气流，这就是大气中的风，可以把一处的气象条件吹到别的地方。风起云涌，原来是亿万分子做无头运动和相互撞击的结果。

大气中的冷热与气压

空气中的分子吸收太阳光或者地面反射的光能后，随机地向各个方向在做极高速度的运动（热运动）。如果没有这些热运动，那么空气的温度将是零下273.15度。那样的话，我们就没法生存。

大气温度为20度时，地面大气的分子热运动平均速度大致是每秒460米。这是吸收了太阳能的结果，这么大的速度也是对这么高温度的反应。当然，我们说温度才20度，比开水沸腾的温度低多了。其实，这个20度，是与水结冰的温度相比得来的，称为摄氏度。

实际上，反映热运动速度的温度，要与热力学温度零度相比。所谓热力学温度零度，是分子再也没有热运动的温度。是零下273.15度。如果与热力学温度零度相比，那么20摄氏度就是293.15度，这个加了273.15的温度，单位是开尔文。293.15开尔文够高的了，所以其数值蛮接近每秒460米的分子热运动速度。

每个分子热运动的方向看似是随机的，不是朝一个固定方向走，平均而言只需移动68纳米（即0.068微米，头发丝直径接近100微米，即在比头发丝直径小一千多倍的距离内），就会与其他分子发生碰撞。在1秒内，每个分子会被其他分子撞击超过10亿次。1秒内空气分子撞击我

们人体的次数，就多得数不清了。当然，这么高的速度，不会把你的皮肤撞出坑来，因为分子太小了。

分子热运动导致空气分子之间相互撞击或者撞击物体而产生的气压，在地面附近是101325帕，大致是每平方米1万千克。成年人皮肤总面积大致是1.5~2平方米。如果是2平方米，那么大气压给身体的全部压力是20吨。这么大的压力，人居然没被压扁。这是因为，我们的肌肉骨骼、体内水分和血液能提供相应的内压，能与大气压力达成平衡，因此不会被压扁。

空气的气压也可以改变。首先，不同高度的空气气压不一样。比如说，海拔30千米的气压不到地面气压的1/100。也可以通过外力将一罐空气压缩成不同大气压。汽车发动机的进气与排气阀关闭后，靠活塞运动可以把汽缸内的空气压缩10倍左右从而气压大致增高10倍左右。飞机在飞行时，正是通过拥有特殊形状和姿态的机翼与空气发生作用，改变上下表面的气压，使下表面气压大，上表面气压小，产生飞起来的升力。

气压变化的传播与烽火台（图3.10）

分子热运动以接力棒形式播送气压变化即播送声音

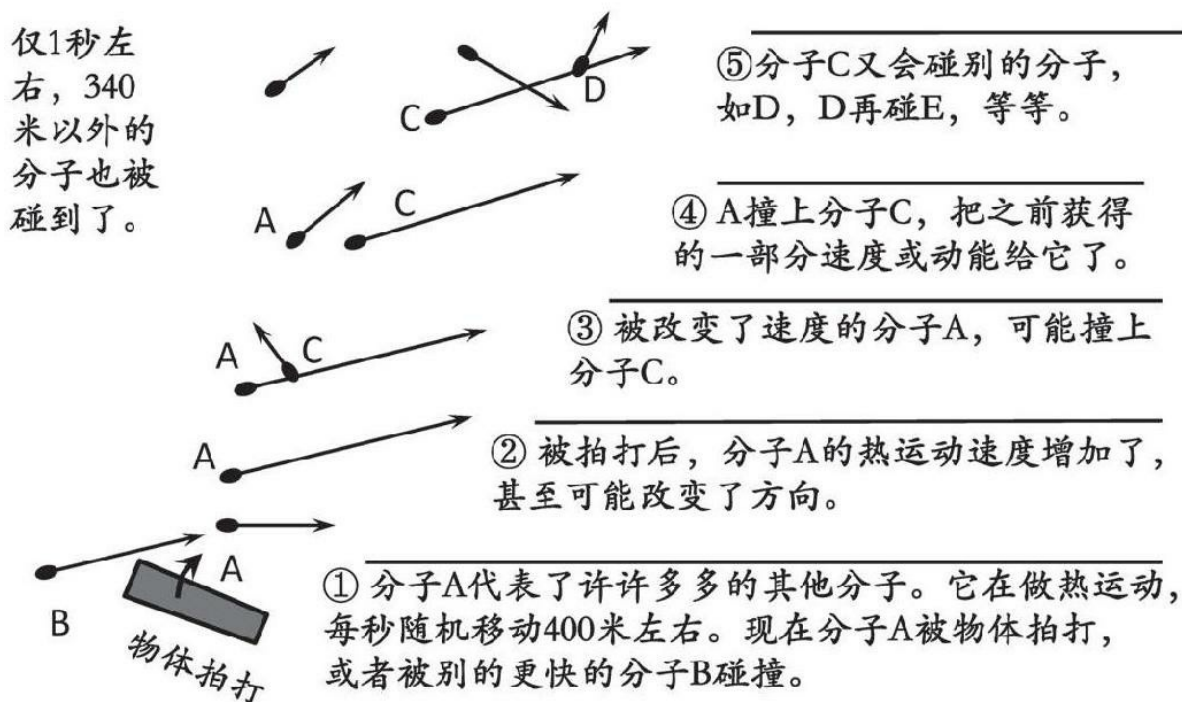


图3.10 声音在大气中的传播

古代战场如果狼烟四起，那是烽火台在传送敌情。我们说话时，物体发出声音时，空气分子最听话。这些分子凭借自身热运动，像田径赛场传接力棒似的，把反映声音的气压变化，以接近烽火台传递敌情的速度传递到远方。那么多混杂的声音我们都能听见，那得有多少空气分子传递声音信息。这里我们先不去理解我们如何能听到能反映不同意思和效果的声音，而是只看看气压变化怎么被送到别的地方，以及以什么样的速度。

我们在这里喊话，一秒内，多远距离的人可以听到？约为340米。就是说，地面空气中声波的传播速度是每秒340米左右，即一小时1224千米左右。我们在这里喊话，340米开外的人在1秒后听到。从北京喊话，长沙那边在一小时后就能听到（如果声音没有衰减到听不见的

话)。

即使是12级台风，每秒也就移动30多米，一小时不到140千米，因此声音传播的速度不可能是气流传送决定的。原来，声音是气压的微小脉动，这种微小脉动能以压力波形式传播，就像水面涟漪是种水位脉动。涟漪水位差传播是靠表面张力拉扯和重力驱赶引起的。但大气中的声音不是这样。

声音是空气中气压忽大忽小的变化，作用在耳膜上引起的。如果没有忽大忽小的变化，而只是缓缓增加，那么听不到声音，只有压迫感。我们听到的音乐，实际上是感受乐器的弦和鼓膜的振动给大气空气扰动引起了气压的忽大忽小的变化。

声波中气压一高一低的变化对应的时间间隔，称为周期。一秒内变化多少次，叫频率。实际上，我们听到的声音包含了各种大小不一的频率的气压脉动。我们的耳朵选择性地听那些可听到的频率对应的声压变化。声波频率越高，声调就越高，否则就越低沉。

古人用烽火台传送关于敌情的信号。我们都熟悉日行千里马的词句，快马一天也不过跑千里左右。两个烽火台相距30里至50里左右，太远了升起的狼烟就不容易看到了。如果一处点起了狼烟，另一处看到并在一分钟即60秒内也点起狼烟，敌情传播的速度就是烽火台之间的距离除以这个时间，每秒250米至416米左右（平均为每秒333米），相当于每小时在900千米至1500千米之间。而马的奔跑速度不会超过每小时100千米。

声音的传播与烽火台传播有点类似，是由分子热运动导致的分子之间碰撞引起的。正是这种碰撞，将一处的气压变化（即声音）以近似等

于分子热运动的速度传播开来。分子热运动起到了传递压力微小变化的接力棒的作用。

由于地面空气的分子热运动的平均速度大致是每秒460米左右，因此声波的传播速度也接近这个速度。由于碰撞过程需要时间，让气压变化也需要时间，于是延迟了一点点分子之间速度的交换时间。因此实际上，地面空气中声音传播速度大致是每秒340米。

这与烽火台用狼烟传递敌情的速度的中间值每秒333米非常接近。这也许是个巧合。

数数大气中的空气分子

这么多地方说到过空气，大多数地方用了笼统的语言，比如分子很多啊，很小啊。现在是时候看看到底有多多，有多小了。

空气分子的平均直径大致为0.4纳米左右（1米的十亿分之一的长度就是1纳米），在24立方米的一间房屋内，大致有600亿亿亿个分子。在地面附近，一立方米的空气大概有1.225千克。其中，按体积比例，有21%左右的氧气分子和78%左右的氮气分子，还有不到1%的其他气体分子。正是因为空气中有氧气分子，我们才能呼吸，空气吸入血液，补充大脑所需要的氧气。

空气分子的平均直径是氧气分子与氮气分子的平均直径。实际上这些分子并不是一个严格意义上的球，只是在相互碰撞时，产生的效应等效于一个个球而已。空气这样的分子小得很难去记它的质量，有一个记分子数目的标准单位，摩尔。1摩尔的气体包含的分子数，是

$6.02214857 \times 10^{23}$ ，这个数称为阿伏伽德罗常数。1摩尔的空气是29克左右，这个质量除以阿伏伽德罗常数，就是空气分子的平均质量，可以想象有多小。地面空气1立方米大致有 2.544×10^{25} 个分子，因此1立方米有42.2摩尔的空气，再乘以每摩尔29克，就有1225克。因此，地面空气的密度大致是每立方米1.225千克。

这么多分子，我们在空气中，根本无法去把空气当作分子来看待。更没法当作原子来看待。万有引力、电磁力、强相互作用力和弱相互作用力虽然是自然界中四种最基本的力，但我们无法用这四种基本力来描述我们所关心的空气行为。好在电磁力的宏观表现就是气压与摩擦力。我们常提到的气压和摩擦力，实际上就是一种电磁力。物体受到的空气升力和阻力，是表面上气压与摩擦力作用的总和。我们感受到的重力，就是万有引力。大多数情况下，我们只与这些总的力打交道。我们恐怕很难感觉到身体的每个分子原子的质量。

煤气炉在烧火时，空气分子与煤气分子通过热运动碰撞引起了所谓的化学反应。烧火会产生二氧化碳。二氧化碳是由两个氧原子和一个碳原子通过化学键（共享电子）结合在一起的分子。由于化学键的结合力，因此拆开一个二氧化碳分子需要添加能量。反过来，如果将一个碳原子与两个氧原子结合成二氧化碳分子，就会释放能量，这个过程就是化学反应。化学反应需要分子或原子之间高速碰撞，才能使它们有足够的机会拆开或搭建化学键。这种碰撞越频繁，速度越大，就越容易完成化学反应。

一个分子不是碰撞一次就能完成化学反应的，有时可能得碰撞数次甚至数十万次才能有一次机会结合成化学键或拆开化学键，完成化学反应。高空的空气太稀薄，分子之间碰撞频率小，发动机的油料很难与

空气发生化学反应。因此，需要增压（一方面增加了密度，另一方面增加了碰撞速度）来保证有足够高频率的碰撞，使化学反应能完成。

4. 疏密有致 胀压有序 雷霆万钧

售票窗口一打开，稀稀拉拉的人群迅速会聚在窗口，挤压在一起。表演一结束，稠密的广场人群迅速撤离。空气受到扰动后，也有这种情形，疏密的变化或者气压的变化导致了压缩与膨胀。如同水波是一种水位高低的变化，声波是一种气压高低的变化。后面的跑得慢，队伍就会拉开，后面的跑得快，队伍就越来越挤。如果后面的使劲往前冲刺，那就撞在一起被推着走。雷电就是这样，因此雷电的响声比声音跑得快。

疏密与胀压 压缩与膨胀（图3.11）

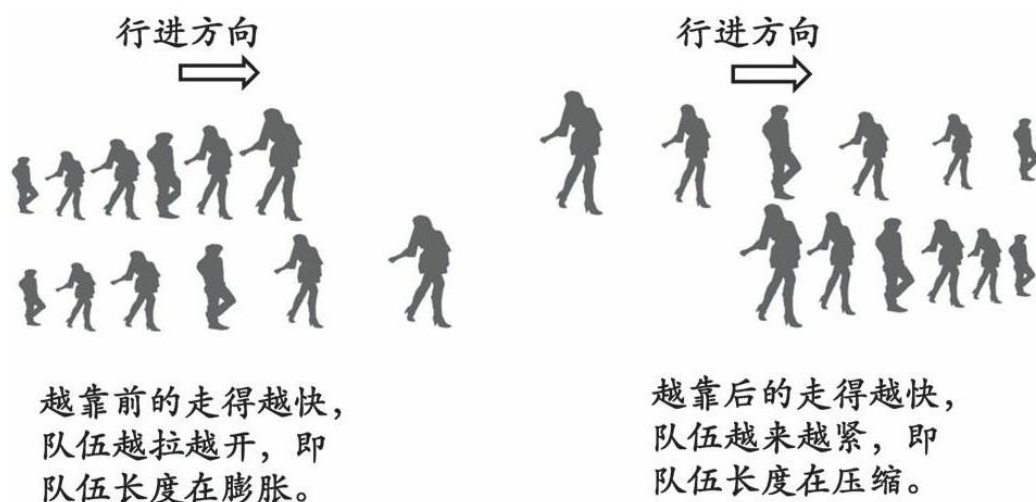


图3.11 队伍的膨胀与压缩

气压的变化可以理解为压缩与膨胀过程。在一个固定位置，气压升高可以理解为压缩（压密了碰撞就多，气压就大），降低可以理解为膨

胀（拉疏了膨胀就少，气压就小）。不断变化有升有降，可以理解为既有压缩也有膨胀。但在空间上看，则不同位置的气压升降是否代表压缩和膨胀，还与传播方向有关。对此，可以看行走的队形来理解。

有若干位同学，在狭窄的山道上行走。假设山道太窄，使相互之间无法超越。每人身高不一，假设越高的同学走得越快。现在排成一队行走。有三种排法。

排法一，产生膨胀队列：让最高的排在队伍前面，往后身高依次减小。这样，越靠前的走得越快，队伍拉得越来越开。于是，队列的长度在膨胀。

排法二，产生压缩队列：让最矮的排在队伍前面，往后身高依次增加。这样，越靠前的走得越慢，越靠后的走得越快。假设各自按自己的速度行走，那么队伍会越来越短，大家挤在一起。于是，队列的长度在压缩。如果每人不顾前面是否有人，继续以原有速度走，那么最终就会挤在一起。挤在一起后，迫使大家调整速度，以一个速度走。

排法三，产生局部膨胀与压缩：不按任何高低顺序排列，在队列各个位置，有高有矮。在这种情况下，假设每人还是按自己的速度行走，那么满足排法一的地方，产生局部膨胀，而满足排法二的地方，产生局部压缩。这种情况往往出现在真实的单车道车流中。每辆车会试图按自己的速度行驶。于是会在不同位置分别产生膨胀和压缩。有的地方几辆车前后挨得很近，有的地方稀稀拉拉，就是这个原因。当然，出现这种局部压缩后，人们会试图调整速度或超车。但还是会出现不同程度的局部压缩与膨胀。

物体在气流中飞行或穿梭，也会引起一些地方的空气压缩，一些地

方的空气膨胀。道理是一样的。有压缩的地方气压增加了，形成了正压。有膨胀的地方气压降低了，形成了负压。于是我们可以用另外一种语言来理解机翼产生升力的现象。从前方到达机翼下方的空气出现了压缩，如同队伍压缩一样。从前方到达机翼上方的空气出现了膨胀。上下表面的气压差就导致了升力。机翼就要做出这样的形状和姿态，使在下方尽可能得到压缩，上方尽可能得到膨胀。

冲击波（图3.12）

物体在空气中运动时，用压力波驱赶前方的空气。压力波传到前方，带来气压差，驱动空气产生气流，就像人被推一下有速度一样。由于压力波每秒几百米，因此前方的空气一般来得及避让。

如果物体的速度超过音速，即超过压力波的传播速度，那么前方空气就得不到物体发出的压力波的提前驱赶。超声速物体到达时，空气来不及避让，被迫在头部附近堆积，形成高温、高压、高密度区。这个区的外部边界也称为冲击波（或激波）。气流从前方穿进冲击波，气压、气温和密度陡然增加。

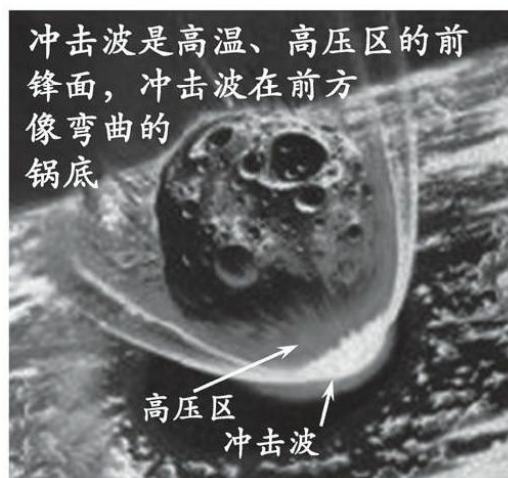


图3.12 流星下落形成的冲击波

头部产生的冲击波会向侧边延伸到离头部较远的地方。飞机产生的这种冲击波会扫到地面，尤其在低空飞行时。原子弹爆炸也会产生冲击波，不过波后的高温、高压来自于爆炸。这种冲击波如果近距离撞击

物体，会因气压和温度突然增高，带来巨大破坏。

雷电声音的速度

说雷电的速度有点模棱两可。电光的速度是每秒30万千米左右，雷电一发生，我们就看到闪电了，几乎没有时间差，因此没什么好说的。值得说的是雷电的声音的传播速度。当然我们会纳闷，不是说了声音的传播速度是每秒340米左右吗？不是，所谓每秒340米的速度，是指声音不太大，气压变化不太强的情况。雷电可是惊雷，响彻云霄，这种巨响怎么也得得上大声音了，引起的气压变化应该很强了，强得可以将空气挤出速度，大得可以让气温暴涨。这样，雷区里外的分子热运动速度以及气流速度就不一样了，到底以哪一边的声音速度传播雷声？得罪哪一边都不行，就像钱塘江大潮的移动速度似的。互不得罪就只能取两边速度的某个中间值。还得注意，此时声波还骑在有速度的气流上，两边声波的速度还得叠加上气流速度，就像飞机的对地速度是空速叠加风速一样。

在对流层积雨云中，水滴、冰晶等相互撞击、摩擦会产生电荷，就像我们人体穿某种衣服时，有时会带电一样。在积雨云顶部往往带正电，下部带负电。这种电位差超过一定程度就会放电，击穿空气，出现闪电，可以让雷电中心的温度接近两万度。这么高的温度，比太阳表面六千度左右的温度还高几倍，难怪发出耀眼的白光。这么高的温度，当然会让气压出现巨大的变化，同时激发强烈气流。巨响的传播，就会受闪电区域内外的气温差和气流速度差影响，会比在安静的空气中转播微小气压变化即普通声音来得更复杂点。

但即使在安静空气中，对小声音的传播的认识一开始就犯错误。

1687年，牛顿这样的大科学家认为，空气在传播声音时不会改变声音经过区域的气温。假定气温不变化，他得到的结果是，地面空气的声音速度约为每秒280米。这比声音的实际速度小了18%左右。事实上，声音传播过程中，气压的微小变化也导致气温的微小变化，牛顿的这个错误直到1816年才由拉普拉斯更正。

气压、气温和气体的密度都跟着发生变化，但总有一种秉性不会变吧。气压变多少，密度不能瞎变。两者的变化量居然遵守某种比例，使在声音传播区，空气的气压与气体的密度的1.4次方的比值，总是保持不变。这个比值就是空气的一种秉性，是一种有序无序程度的衡量，物理学上称之为熵。熵就是衡量有序无序程度的量。摩擦和热量的传递会让熵增加，即导致越来越无序。

可是，声音传播过程极快，气压只有微小变化的声音，在这种极快过程中，摩擦和热量传递来不及起作用，因此这个比值是不变的（即熵不变）。拉普拉斯正是通过假定这个比值不变（物理学上称为熵不变，也称为过程可逆，即声音过后空气恢复到原有状态），才得到了正确的音速。

响雷也是一种声音，但与一般的声音有区别。一般声音传播过程中，空气的气压变化是极其微小的，但响雷引起的气压变化很大。厄恩肖（Earnshaw）在1851年，发现雷电引起的响声，比普通声速快，即雷电以超声速传播。

不难想象他如何测量雷电响声的传播速度的。我们都熟悉有雷电时，先看到闪电，数秒后才听到响声。原来，闪电出现那一刻，就发出了响声，闪电的光速每秒30万千米，因此传到眼前需要的时间可以忽略不计。闪电出现的位置与人有一定的距离，将这个距离除以闪电后听到

声音所需要的时间，就得到了雷电响声的传播速度。

由于声音的传播速度是分子热运动引起的，因此空气温度越高分子热运动速度就越大，从而声音速度就越大。所谓每秒340米的声速，实际上为温度是15摄氏度时的声音速度。当温度是27摄氏度时，声音的速度是每秒347米左右。

雷电将空气击出极高的温度（近两万度）和气压，压力波向外传播时，高温区温度高，压力波传播得更快。温度按开尔文为单位每翻两番（四倍），声音传播速度就翻一番。因此，雷电高温区的声音比静止大气中的声音传播更快。

高温区的压力波传播更快，套在外面的静止大气的压力波传播更慢，结果像长江后浪推前浪一样，高温区的压力波与静止大气的压力波挤在一起，形成类似于钱塘江大潮的潮头或者原子弹的冲击波一样，以两侧声音速度的某个中间值传播。

这个中间值比静止空气中的声速显然大，因此雷电的响声相比于静止空气，是以超声速传播。实际上闪电高温区还有膨胀带来的气流，因此声音的速度还得叠加上气流速度，就像飞机对地速度是空速叠加风速一样。这使雷电响声传播速度更快。

当然，传播足够远后，气温已经降下来了，因此传播速度会越来越接近静止大气中的声速。

5. 火旋风 空气中的旋涡 昆虫与鸟的旋涡

小旋涡几乎无处不在，只是混在一起看不出来。气流方向交叉或者平行气流的速度差都等效于一种旋转。旋涡也会弱肉强食，大的吞小

的。相互吞并就会形成更大的旋涡。旋转让人觉得美丽，如波浪云中的旋涡和卷起树叶的小尘卷，也包括小蜜蜂和蝴蝶的旋涡。也会令人恐惧，如火旋风、台风、龙卷风和沙尘暴旋涡。大气旋涡可以带来雨水，植物成长它们功不可没；偶尔也带来灾难，例如火旋风与龙卷风。火旋风拉着长长的火焰，被举高的火种更容易飘到远方，加剧火灾的蔓延。旋转不仅让周线上有风速，也让涡心降压，降低气压后，使涡心浮力更大，产生沿涡心的上升气流，将周边的空气卷入。周边空气如果拥有不均匀的小旋涡，被带入涡心后就加强旋转速度。

火旋风 蓝旋风（图3.13）

火旋风并不常见，偶尔在森林火灾、城市建筑物火灾和居家火灾中看到。旋转的火焰在离心力的作用下，降低火焰中心轴中的气压和密度，使得火焰中心有较大的浮力，火焰就比没有旋转的火焰升得高。这会把火种带到很高的高空，加剧火灾的蔓延。

虽然火旋风只是随机地偶尔才出现，人们还是好奇火旋风是怎么产生的，即火焰为何会旋转起来。对于一个特定的被观察到的火旋风，可能来自于当地的气流本身在旋转。密集森林、山丘和建筑物的背风面，气流在那里折转，容易出现拐弯涡。当这种旋涡中心与火焰重合时，火旋风就形成了。在有旋涡的着火区，大气旋转更容易把周围的氧气从底部带进火焰，使本来缺氧的燃烧更强烈了。

火旋风也可以由火焰自发形成。大面积燃烧可能导致一堵堵火墙出现，火墙来自于那些尚未烧干净的火源。火焰区由于浮力，气流会上升，这需要水平气流来补偿这种上升后留下的空隙。火墙会阻挡这种水平气流，但火墙之间的缝隙有利于水平气流通过。如果缝隙方向使气流

方向交叉，就会引起旋转。

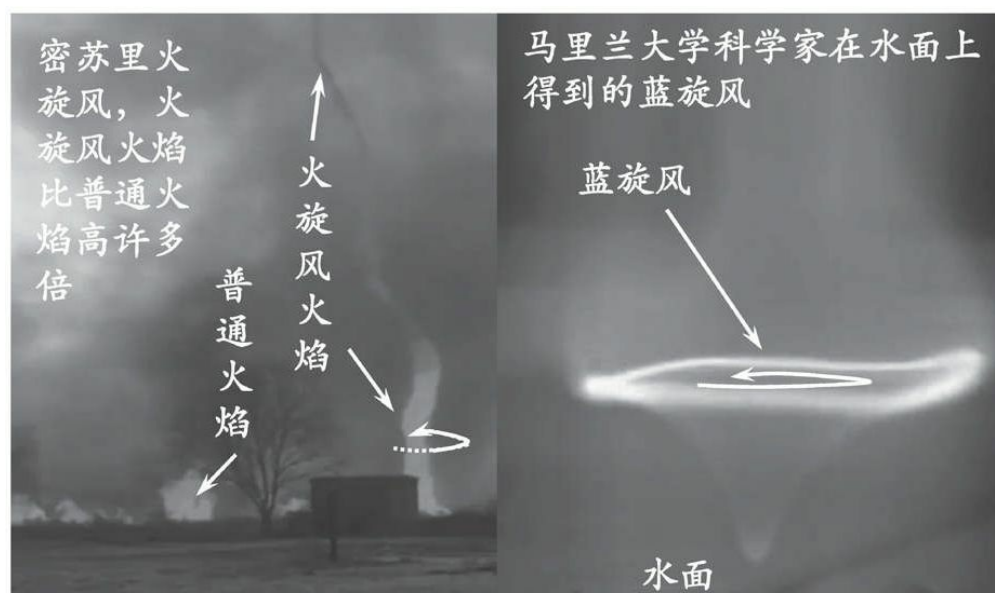


图3.13 火旋风与蓝旋风

炽热的火焰，包括火旋风，一般是红色的，这是因为烧得不是很充分。我们的煤气炉点火后，如果火焰是红色的，表明燃烧不充分，质量不好。正常情况应该是蓝色火焰。

马里兰大学的肖、格尔纳和奥兰通过在水面上生成火焰，发现了蓝旋风现象，即旋转的火焰是蓝色的。这种蓝旋风把燃料燃烧干净了，没有污染物排放，因此呈蓝色。

在地面燃烧时，火焰中气流上升需要带动地面附近的气流流向火焰。由于地面的摩擦，这种向中心的流动受到阻碍。于是，火焰中心的氧气得不到有效补充，缺氧燃烧就成了红色的火旋风。

但如果是水面，那么流向火焰中心的气流在水面滑动就比地面顺畅多了。这使火焰能高速旋转，新鲜空气补充及时，从而使燃烧更干净，

火焰就成了蓝色的。

这一重要发现对清洁燃烧以及海面溢出的油污清除有帮助。

大气旋涡各有姿色（图3.14）

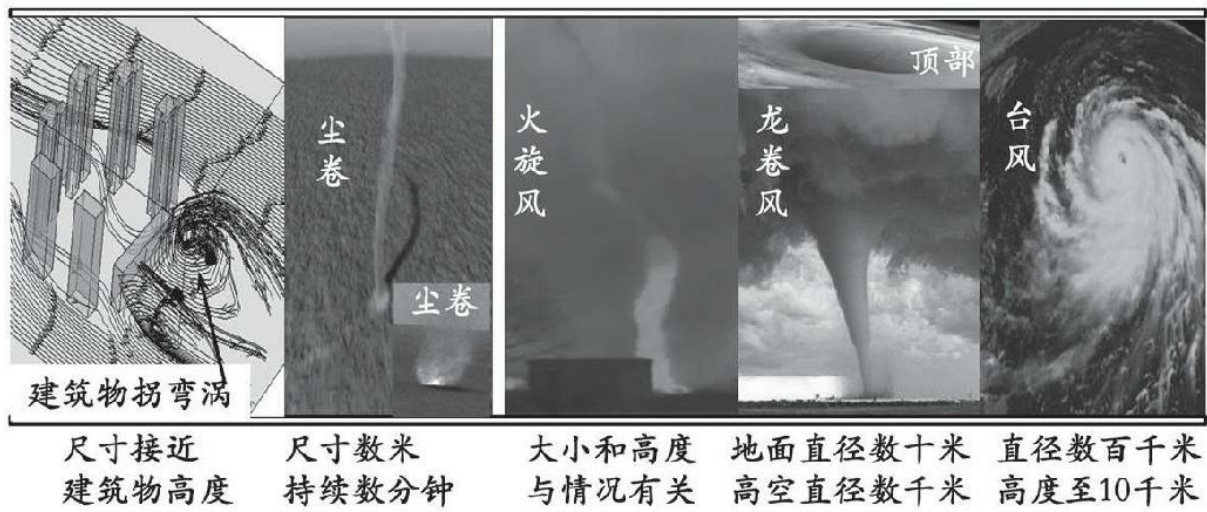


图3.14 各种旋转气流

我们常见的大气中的旋涡，有尘卷、火旋风、龙卷风以及台风。

我们可能在草地、沙漠和高楼边见过尘卷。尤其在高楼边上，大风吹来，地面树叶和尘土突然卷起，围着一根竖轴打转转。有时持续数秒，有的间或地持续更长时间。民间有的称之为鬼旋风，埃及称之为幽灵风。在世界各地称呼较多。比如说在美国南部称为舞旋风。尘卷，顾名思义，就是把尘土卷起旋转的旋涡。

从高楼、树木以及山丘的侧面过来的风，转到背风面时，拐弯过程容易导致拐弯涡，这种旋涡可能引起我们见到的尘卷。但标准的解释是地面局部临时产生的热气团在浮力作用下上升时形成的。上升过程把周

边的空气卷吸进来。周围空气或多或少带有一些杂乱的小旋涡。如果小旋涡分布不均匀，汇聚到上升气流中间后，这些小旋涡就合并成可卷起尘土的大旋涡，即尘卷。大旋涡形成后，这种卷吸周边小旋涡的能力就更强，使尘卷可能持续一些时间，并可能在风中往一个方向移动（被风吹动，或者由于一侧的卷吸力大于另一侧）。

火旋风由于自地面火焰产生，因此主要是地面因素主导旋转，带动旋转后，越高的地方这种因素越弱。因此火旋风底部粗，高处较细。

龙卷风则顶端更粗，底部较细，因为这样的龙卷风主要来源于高空超级雷暴区本身的中尺度气旋，带着下方的大气旋转。

令人畏惧的台风（图3.15）

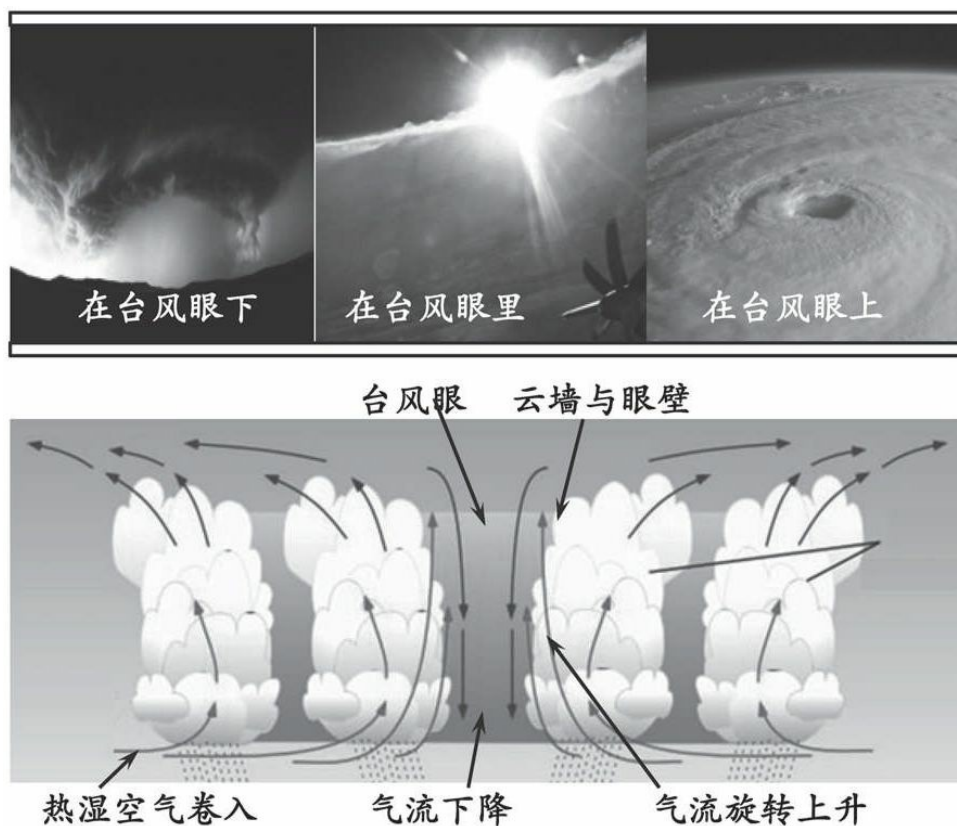


图3.15 台风模型与台风眼

我们所指的台风，在西太平洋地区，如中国沿海一带称为台风，在东太平洋地区如美国称为飓风。

台风虽然是旋转的，但由于尺寸巨大且其拥有特有的结构，地面上的人是看不到旋转的。只有在卫星或高空飞机上才能看到旋转现象。

海洋温度增高后，蒸发的水蒸气升入高空，与空气混合形成潮湿气体。更靠近赤道的海面受地球自转影响，由西向东的速度更高。更靠近北极时，地球自转效应更弱。因此，热蒸汽上升后，其带动的指向东方的气流速度沿着纬度方向有差异，越靠近赤道越大。这种速度差有搓出旋涡的能力，与波浪云滚出旋涡道理是类似的。这种速度差被沿经度方向的风吹到某区域集中后，就得到加强，很容易旋转起来，形成台风。由于是靠近赤道一侧速度比靠近两极的速度大，因此，这种速度差搓出的旋涡，从卫星上看，在北半球生成的显然是逆时针旋转，在南半球是顺时针旋转。

由大面积海洋蒸发引起的台风气旋，直径可达数百千米。旋转造成旋转中心出现低气压，有可能形成直径数十千米量级的台风眼。台风眼中心可能风平浪静，但台风眼的边缘（台风眼壁）风速极高，可达每小时一百千米的量级。远离台风眼壁的气流旋转速度则会降下来。

台风眼壁的转速可达每小时一圈的量级。据说，台风眼壁外围可形成环形云墙，云墙围成的区域里面相对安静，云墙附近风速极高。按此台风模型，台风中心附近最高风速，应是指台风眼壁处的风速。某处台风的成因、尺度、风速、台风眼壁的大小、形状，是否有云墙等，还要看具体情况。

虽然我们一听到台风就害怕，但台风带来雨水，是地球生态的重要组成部分。

昆虫、鸟与飞机激起的空中旋涡（图3.16）

那些在空中飞行的物体，也会激发出旋涡。小的昆虫的旋涡在空中封闭成一个个涡环。鸟翅膀拍动较慢时，旋涡如同机翼的翼尖涡，是一种气流交叉产生的翻转旋涡。拍动较快时，就和昆虫的差不多，也会形成闭环。科学家证明了一个定理，说旋涡要么被拖到无穷远，要么止于陆地或水面，要么形成闭环，不可能在某点断掉。这就是所谓的旋转代表的强度（涡强）守恒原理的一种体现。不能就质量可以守恒，能量可以守恒，惯性可以不变嘛，旋转的涡也得守恒。

蝴蝶的翅膀采用打开合拢方式。打开时，空气试图填满留下的空隙，但从翅膀外侧进来时需要拐很大的弯，于是急急地滚出一对拐弯涡，每侧翅膀里各有一个。合拢时又会把这对涡吐出去。

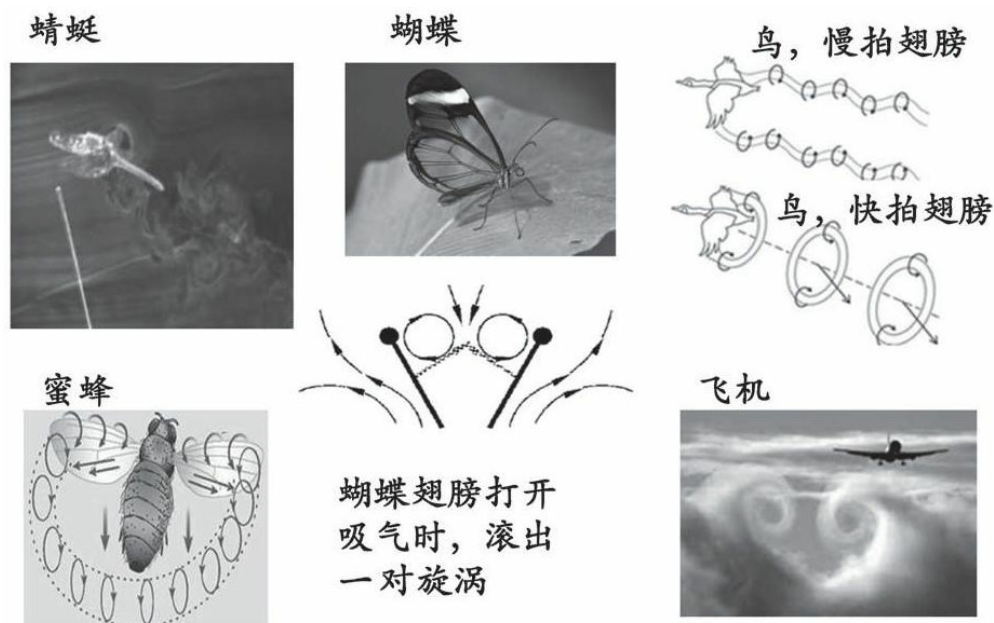


图3.16 生物空气动力学家发表的昆虫和鸟的旋涡

蝴蝶不仅本身美丽，旋涡也美丽，而且很安静。不信，就试试能不能听到蝴蝶飞行时的声音，可比蚊子安静多了。

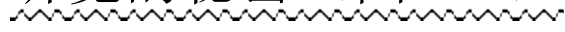
3.2 听懂世界

大自然的美丽靠我们的双眼去欣赏，因为我们眼睛有视觉功能。除了视觉外，我们还拥有听觉、嗅觉、味觉和触觉，其中听觉接受的是空气中传播的声音。我们从声音中也能听懂一些现象。有的是噪声，甚至像爆炸一样的噪声。有的听不见，但会压迫耳朵。有的是美妙的音乐。原来，耳朵接受声音信号后是通过耳膜受迫振动和共振来感觉到声音。于是，不仅是声音的强度，还有声波的频率以及产生声音的节奏和叠加方式，决定了声音是烦人还是悦耳。于是，不仅从视觉可以看出什么，还能从声音听出什么。由于声音的频率与振动物体的长度有关，因此不难理解高大的男人声音低沉，娇小的女人声音尖细。

1. 听不见与听得见的声音 蝴蝶与大象

我们熟悉大智若愚和大象无形这样的哲学语言，其实也有大音希声这样的真实现象。我们还可能听说过陶斯之声的故事：一些人声称听到了一种怪怪的声音，其他人却听不到。这可能吗？我们到底是怎么“听到”声音的？听到声音其实就是空气中气压的变化引起耳膜共振被听觉神经感觉出来的。如果气压的变化与耳膜不同拍，那么任凭怎么呐喊，就是听不到。

听觉的秘密（图3.17）



说话的声音、音乐、虫鸣鸟叫、运动物体发出的噪声等，引起了空气中气压随时间的脉动。气压一高一低的变化对应的时间间隔，称为周

期。一秒内变化多少次，叫频率。声音中气压脉动一般包含了许多频率的气压脉动。可以看成数目众多的频率的气压变化叠加在一起。

我们的身体暴露在大气中，却感觉不到这种气压的微小变化。不知是否有怪人怪物，真的能感知这种气压变化，当作声音来理解。

但耳朵不一样，耳朵就是进化出来的，可以感知声音的器官。耳膜接收到声音信号，在声波气压变化的作用下，振动起来。这种振动被听骨放大，再刺激听觉神经，就能形成听觉。不难理解，耳膜也是一种振动机构，不是什么频率都能振动得出来，只有与声音频率合拍的那些振动才能被听到。高了低了都不会有反应。这就像荡秋千，你推得太快或者太慢，都不会让秋千荡起来。

事实上，由于耳膜的这种限制，我们的耳朵能听得见的声波的频率大致介于20赫兹到20000赫兹之间（这个范围也称为听觉频率）。也许耳朵因为受到过什么外力作用变形后，听觉频率范围会有些变化。或者某些人的耳膜尺寸不一样，听觉频率范围就不完全不一样。如同振动的弦，长了短了其发出的声音就不一样了，因为振动频率不一样。

一般人对20赫兹以下的声波（次声波）和20000赫兹以上的声波（超声波），是难以听到的。如果声音尤其是噪声包含了从小到大各种各样的频率，那么我们的耳朵只是听到了那些频率介于20赫兹到20000赫兹之间的部分。

如果气压用1秒的时间单调地增大了1000帕，这个频率就只有1赫兹，因此我们听不到相应的声音，只是有压耳的感觉。坐飞机在起飞着陆阶段，由于不同高度气压不同，就有这种压耳感觉。雷电的频率大概是50赫兹，所以听得到。人吹口哨是1000赫兹左右，因此也听得到。但

狗吹口哨时，频率太高，人就听不到了。

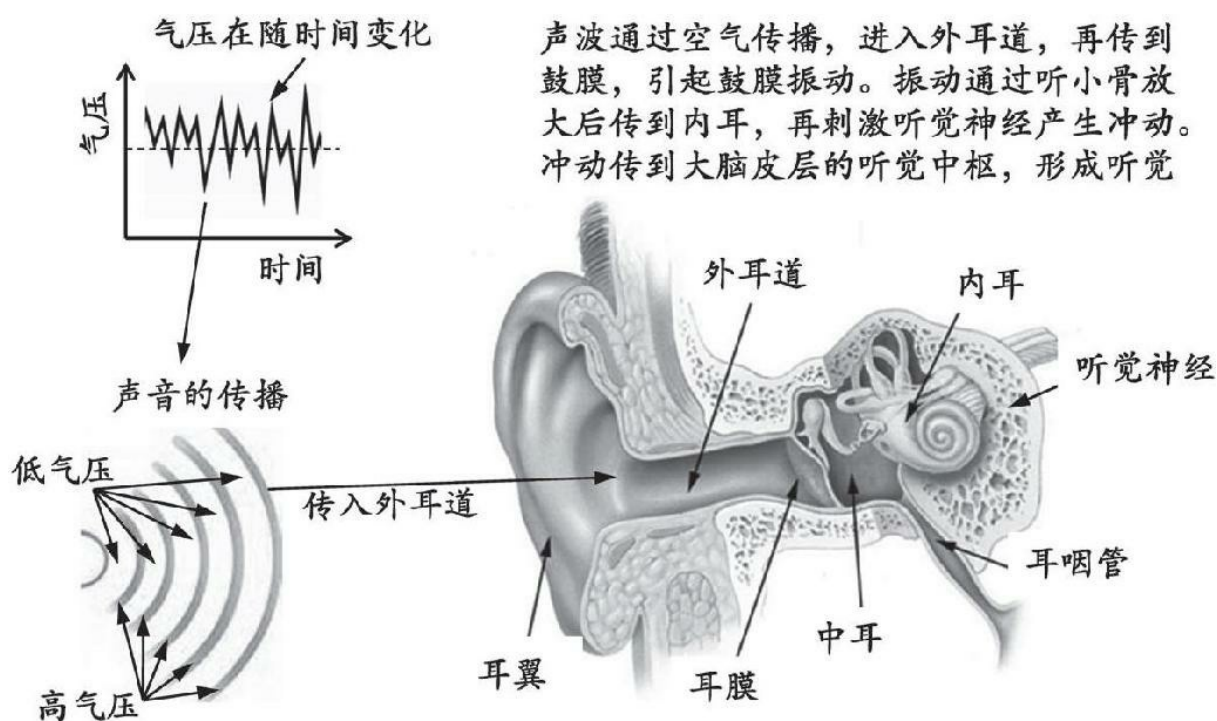


图3.17 听觉产生的原理

陶斯之声 蝴蝶与大象

虽然如此，少数人的耳朵可能听到20赫兹以下某些频率范围的声压变化，也许这就是陶斯之声的秘密所在。

蝴蝶每秒振翅5~10次，即翅膀扇动的频率在5赫兹到10赫兹之间。人们往往以此解释蝴蝶翅膀扇动的声音为何听不见，却看得清翅膀扇动。

唐朝诗人杜甫的《蝴蝶》有“穿花蛱蝶深深见，点水蜻蜓款款飞”这样的诗句。可见蝴蝶在穿越花丛时，可以让你看得清清楚楚，包括扇动

的翅膀。首先，人们把看得清翅膀归结于我们视觉的能力。某认知神经科学家说，经过训练的飞机驾驶员一秒内可以识别255帧图像（即他们能看到只出现1/255秒的飞机图像）。对于普通人，除极少数人在一秒内只能识别不到24帧外，绝大部分人能识别24帧，半数以上的人大概是45帧，大约四分之一的人每秒可识别超过60帧。因此，电影每秒走24帧不同图像，造成人物活动的感觉。蝴蝶扇动翅膀频率远低于24帧，因此在扇动过程中可以看得清翅膀。

人们往往把蝴蝶翅膀引起的声波频率等同于翅膀扇动频率（5到10赫兹，未达到听觉频率下限），以此来解释为何听不到翅膀扇动的声音。然而，听不到蝴蝶声音的道理并不这么简单。

按照上面这种简单解释，那并没有拍动机翼的飞机，按理频率为0，就不应该有声音。但实际上能听得到飞行时机翼发出的噪声。其实，这种噪声与机翼本身是否拍动无关，而是来自于空气流过机翼时产生的空气摩擦旋涡与湍流引起的气压高频变化。湍流脉动包含气压脉动，频率最高可达10万赫兹甚至以上，旋涡引起的气压变化频率可覆盖1000赫兹到10万赫兹。

如果是巨大的人造蝴蝶，用5到10赫兹的频率扇动巨大的翅膀，还是有声音的。真实的蝴蝶没声音，是因为翅膀很小，翅膀扇动幅度小，扇动引起的气流与气压变化并不产生机翼那样的高频变化。气压变化几乎与翅膀动作同步，引起的气压变化频率几乎就是翅膀扇动的频率。因此，蝴蝶扇动翅膀引起的声音的频率在5赫兹到10赫兹之间，是人耳听不见的。除此之外，蝴蝶扇动翅膀引起的气压变化量也很小。所谓的蝴蝶效应另有所指，正常情况下蝴蝶展翅掀起不了什么风浪。

翅膀扇动最慢的昆虫是一种黄凤蝶，翅膀扇动频率仅有5赫兹（每

分钟300次），而大多数蝴蝶的翅膀每分钟扇动460～636次，即每秒低于11次。据说大象能发出并听到1赫兹的声波。也许大象能听得懂蝴蝶。由于人的生理限制，还不知道有多少美妙的声音我们听不到。

但不是所有昆虫都像蝴蝶那样。蚊子的翅膀频率就可以高达100赫兹以上。世界上翅膀扑动最快的是一种俗称小黑蚊的蠓，它每秒可拍动翅膀1046次，即每分钟翅膀扇动次数高达近62800次。蚊子扇动翅膀的声音当然能听到，因为频率在听觉范围内。但是，它们的声音好听吗？

2. 烦人的噪声 高速物体的音爆

噪声的频率怪异混杂，引起耳膜乱抖，听起来当然不舒服。这如同一只小动物被人拽住乱抖很不舒服一样。气压变化的有序或混乱程度决定了包含什么频率，气压变化的幅度决定了声音的强度，即声音有多响和多大。太响，也会引起不舒服。高速列车进入隧道后，出口可能出现爆炸声，从高空掠过的超声速飞机也会引起突然的爆炸声，这就是响声太大的音爆。气压变化的幅度虽然与大气压相比不算大，但与我们听觉能感知的最低气压相比却很大，如果来得突然，就会是一声巨响。

声音的强度 对数的作用（图3.18）

我们听觉能听到的最低气压变化是0.00002帕，即每平方米2毫克左右。但我们说话引起的声音，在1米距离，比这大1000倍左右。距繁忙的道路5米处，车水马龙的声音比这大1万倍左右。听起来让人不舒服的声音比这大100万倍左右，压得耳朵疼的声音比这大600万倍左右。因此，以帕来标记声音的强度，数值大小跨度太大，很不方便记忆。

于是采用对数记忆法。以10为基准的对数为例，10的对数是1，100的对数是2，1000的对数是3，10万的对数是5，100万的对数是6，如此类推。即如果一个数的1后面有几个0，那么对数就是这些0的个数。

因此，如果说话的声音气压变化量是听觉下限的1000倍，用了对数后就只是3倍。难听的声音是听觉下限的100万倍，用了对数就是6倍。这样记忆起来就方便多了。但这样未免显得太小了，于是人们使用了分贝来表示声音强度。分贝就是气压变化幅度相对于听觉下限（0.00002帕）的比值取对数后，再乘以20得到的数。例如，比听觉下限强度大1000倍的声音（对数是3），就是3乘以20，等于60分贝。这就是我们说话的声音的强度，用分贝表示就是60分贝。一般要求噪声不要超过80分贝（听觉下限的10000倍）。

声音引起的气压变化会随距离衰减，因为气压变化也是能量，以球面形式传播，即使只为了保证能量守恒（还不考虑摩擦衰减），也要衰减。在自由空间，距离每增加一倍，声音强度降低6分贝，即气压变化与离开声源的距离成反比。

因此，说一个声音源发出的声音强度有多大，一定要说在什么距离是多大。不说距离，毫无意义。比如，我们说话的声音，1米距离听到的大致是60分贝，1千米以外的人肯定听不到了。说我们呼吸的声音大概是25分贝左右，那是指1米的距离听到的声音大小。我们常说雷声有100分贝，单这样说就很不准确，因为没有指明到底离闪电有多远来测这个强度。北京听到一个雷是100分贝，巴黎就听不到这个雷了。

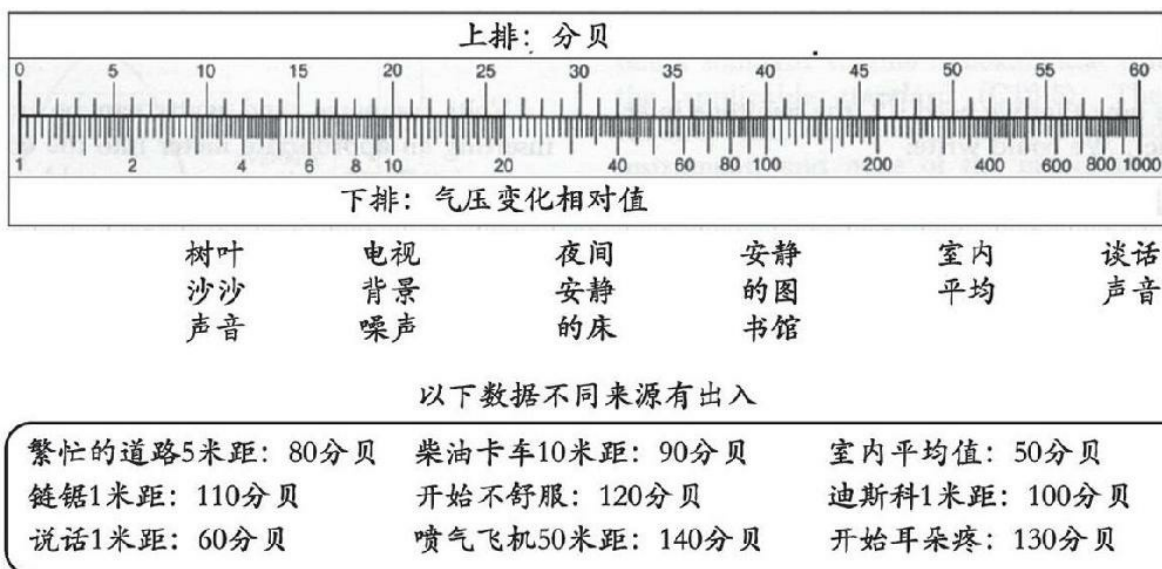


图3.18 不同噪声源的强度

声音强度大并不就一定就听得到，还与气压变化的频率有关。前面已经提到，气压变化的频率在20赫兹以下，或2万赫兹以上，人就听不到了，这是指的强度不太大的声音。其实，如果声音强度很大，那么低于20赫兹的声音也可能听得到。超出20~20000赫兹这个听觉频率范围，如果气压变化很大而听不到时，耳朵会有压痛感。

声音强度太大，会损害听觉，但也与暴露在声音中的时间有关。82分贝的环境下，暴露16小时就会损害听觉。85分贝（气压变化0.36帕）时降低到8小时，94分贝（气压变化1帕）时降低到1小时。100分贝（气压变化2帕）时为15分钟。110分贝（气压变化6帕左右）时为1分钟左右。115分贝（气压变化11帕左右）时为30秒。

隧道出口的爆炸声（图3.19）

高速列车时速可以达到300千米（即地面声速1224千米的0.24倍）

甚至以上。由于速度小于音速，因此列车飞驰时，引起的声波会提前到达前方。列车不但将头部的空气推着走，而且前方的一部分空气会绕到车身边上。由于列车头撞击，车头最前端的气压会比大气压稍微高一点。由于受到猛烈撞击的空气从侧面快速避让，车头与第一节车厢的连接部位的气压可能稍微低于大气压（比如说拐弯时离心力降低了气压）。其他车厢边上的气压则接近大气压，因此我们乘坐高速列车在开放空间飞驰时，耳朵不会有明显气压变化带来的不适。

如果高速列车进入隧道，情况就不一样了。这相当于列车要在隧道中顶着空气往前推。这与汽车发动机活塞问题十分类似。把隧道看成汽缸，把火车头看成活塞，这相当于活塞在汽缸中冲进。活塞向汽缸顶盖方向运动时，会产生压力波将汽缸内空气压缩。

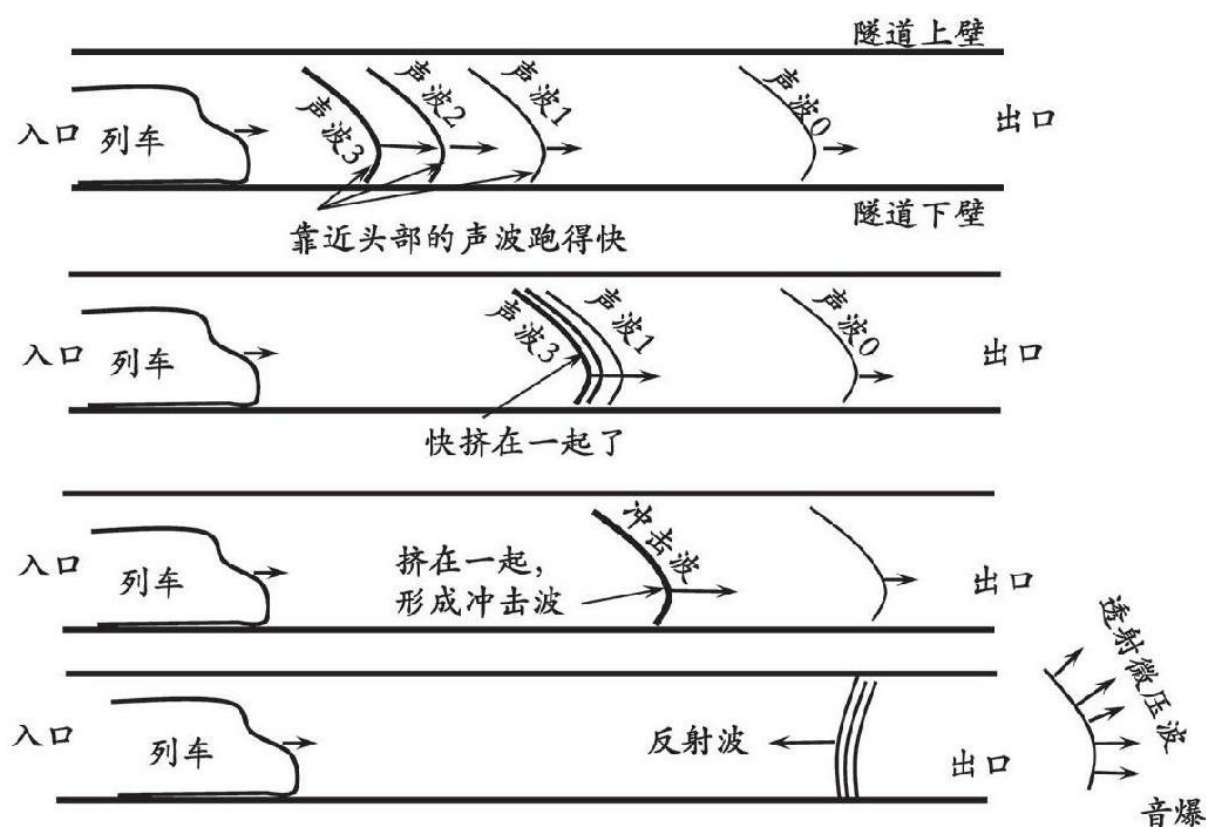


图3.19 列车进入隧道产生的冲击波与出口微压波

越靠近车头的空气受到火车的挤压作用越强，因此压力波就越强，温度越高，向车头前方传播速度越大。因此，车头附近的压力波会试图追趕上车头“远”前方的压力波，最后堆积成一个压力变化很强的冲击波，也称为激波。

冲击波的运动速度比其前方的声波还快，尤其会比火车提前到达隧道出口。冲击波遇到出口，与外界大气相遇，大气迅速调节自己的气压，迎接冲击波的到来。

大气的气压比冲击波的气压低，因此接触时会相互交换一下气压。冲击波下游的气压减少了一点后，对应的气压变化也代表了一种压力波，于是向管道内传播。由于向里传播时是降压过程，因此是膨胀波。冲到外面的冲击波气压减弱了一些，在出口外的大气中传播，“头衔”降了一级，称为微压波了，谁要你没有在隧道中那么强了呢？

虽然如此，这种微压波比普通声音强多了。它在开放大气中进一步传播，周边居民能感受到这种冲击波带来的噪声。由于微压前方的气压为大气压，而微压波里的气压高一些，因此在这种微压波的作用下，耳朵在极短时间内会感受到很强的气压变化，即我们会听到爆炸的声音，即音爆。

列车上的感觉（图3.20）

在车厢里的乘客也会感受到气压变化。相比于环境大气压，乘客感觉的气压变化量可以达到1000帕甚至更高，相当于每平方米100千克甚至更高。对于面积约90平方毫米的耳膜，感受到的压力变化是9克甚至更高。这么大的变化量是在一定时间内完成的，对应的频率并不高，因

此耳朵听不到大的声音，只是感觉有压痛。

在过隧道时，车厢内气压变化的原因较为复杂但可理解。列车进入隧道时，车头产生的冲击波传到出口后，一部分会变成减弱的冲击波（称为微压波）传到大气中，一部分会反射成膨胀波向隧道内回传。膨胀波掠过一点时，使当地原有气压降低。膨胀波进一步在火车头反射成向前传播的膨胀波，余下的一部分经过隧道缝隙传到入口，在入口又反射成向隧道内传播的（很弱的）冲击波。

假设隧道长是1千米。时速300千米的列车用12秒穿越隧道。冲击波与反射膨胀波传播速度接近或稍微超过声速，即时速接近1200千米，穿越隧道只需要3秒左右。因此，列车过隧道时，会数次遇到冲击波或膨胀波。由于冲击波增压、膨胀波降压，因此车厢里一会儿气压增加，一会儿气压减小。

车头进入隧道引起冲击波，车尾进入隧道则引起膨胀波向隧道内传播，因为车头进入与车尾进入对隧道内空气的影响方式是相反的。

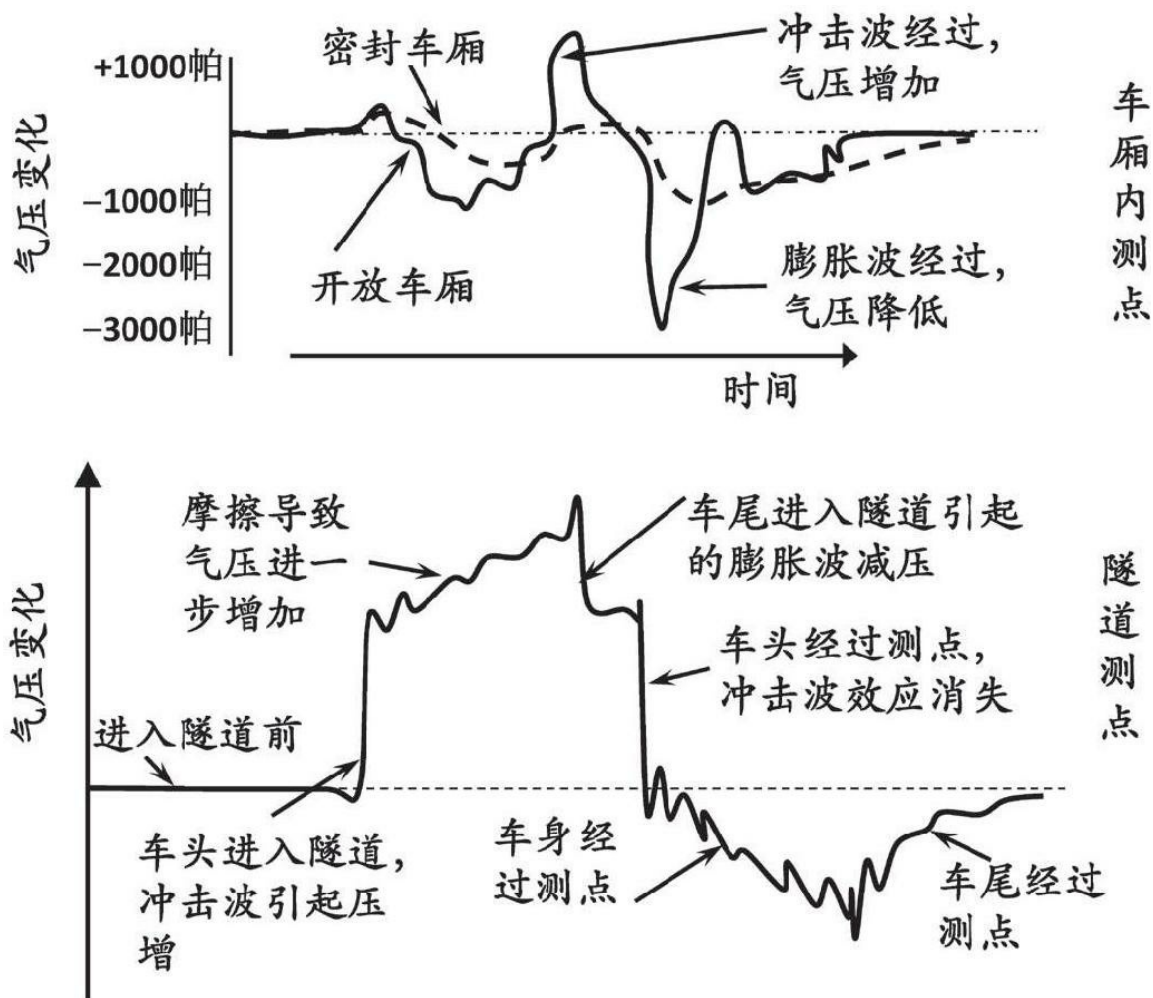


图3.20 车厢内一点和隧道内一点的气压变化

飞机的音爆（图3.21、图3.22）

飞机在高空飞行时，乘客的耳朵会习惯高空飞行时舱内的气压。着落下降时，由于低空气压高些，于是就会有一个气压越来越高的过程。这使一部分乘客的耳朵有压迫感。飞机本身发出的声音，除了影响乘客，地面也会受影响，尤其超声速飞机带来的音爆。

超声速飞机头部有一道冲击波，会传到地面，使地面气压突然增

加。飞机从机身中段某位置开始，横截面积收缩，出现与冲击波相反的效应，即引起膨胀波，使气压降低。这也会传到地面，使被冲击波抬高了的气压又下降。这种膨胀带来的降压往往比第一道冲击波的增压幅度更大，这需要再次产生一道冲击波来将气压提升到大气压。因此，超声速飞机从头顶飞过时，会听到两次音爆，前后相隔0.2秒左右。这种气压先增加，接着线性下降，再增加的现象，放在时间轴上，颇像N字形，因此也称为N型音爆。

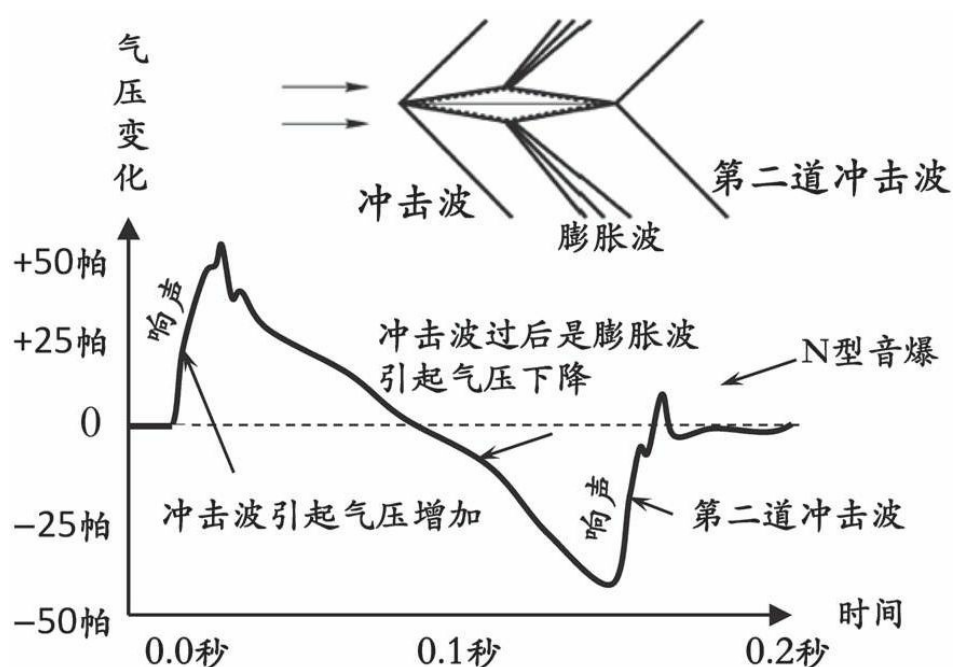


图3.21 超音速飞机冲击波和膨胀波引起N型音爆

在24千米高空以3马赫数飞行的黑鸟（SR-71）产生的音爆在地面引起的声音是43帕。航天飞机在18千米高空时，已经降至1.5马赫，地面声音是60帕。在16千米以2马赫飞行的协和超声速客机，地面听到的声音是93帕，超过了80分贝。由于超声速飞机音爆扰民，它难以作为民航客机推广。



图3.22 飞机产生的冲击波

一些马赫数小于1但接近1的飞机，在机身中段也可能产生冲击波，从头顶飞过时，也能听到一声音爆。这种飞机如果做超低空飞行，那么地面被冲击波引起的气压变化足以造成破坏。

电影《空中决战》中，幻影2000战斗机在袭击东非机场的时候，以接近音速掠地飞行，产生冲击波，将地面建筑物玻璃都震碎了。

说话、音乐、噪声等，声音的强度有可能相同，噪声烦人，但音乐好听，因为里面包含的频率和节奏不一样。现在让我们听听什么是好听

的声音。

3. 动听的声音 音乐与频率

噪声烦人，音爆扰人，前者因为频率混杂，后者因为突发巨响。声音是否好听，与构成声音的频率的组合有关。水面产生的涟漪和涌浪，以及天空中的波浪云，按节奏出现在不同位置，从而形态优美、赏心悦目。同理，具有和谐频率的声波按时间叠加，可能产生动听悦耳的音乐。乐器依靠琴弦、气腔或鼓膜的和谐振动，激发可用于和谐叠加的声波频率，各段谐波通过时间长短控制和恰逢其时的叠加获得有动感的节奏，从而可以谱写出优美的旋律。一万小时成才定律用了包括柏林音乐厅在内的音乐家成才的历史作为了主要数据，这说明让乐器能弹出优美动听的音乐需要成千上万小时的练习。音乐是需要技术且超越技术的艺术，这里只能简要地涉及其中与频率相关的一点点知识。

琴弦的振动与声音（图3.23、图3.24）

我们能轻易地把一根尼龙绳抖动起来，产生像波浪一样的形态。尼龙绳有张力，你双手握住一小段，往两边扯，就会感到吃力，扯得越长越吃力。尼龙绳有重量，平均每一米有一个重量（称为线重），线重越小越容易抖动起来。越长，总的重量越大，也越难抖动。

小提琴、吉他和钢琴上的琴弦，两端固定，弹奏起来也会像抖动的尼龙绳一样舞动起来。如同尼龙绳的张力越大越容易振动而线重和长度越大越难振动一样，琴弦的振动频率也与这些因素有关，且在一定的条件下最基础的频率（即最小的频率，也称为一次谐波）满足梅森定理，该定理指出，最基础的频率（基频）与弦长成反比。除了最基础的频率

外，还有可能产生频率倍增，在不改变弦长情况下形成高频谐波（称为二次谐波、三次谐波等，其频率分别是基频的两倍、三倍等）。

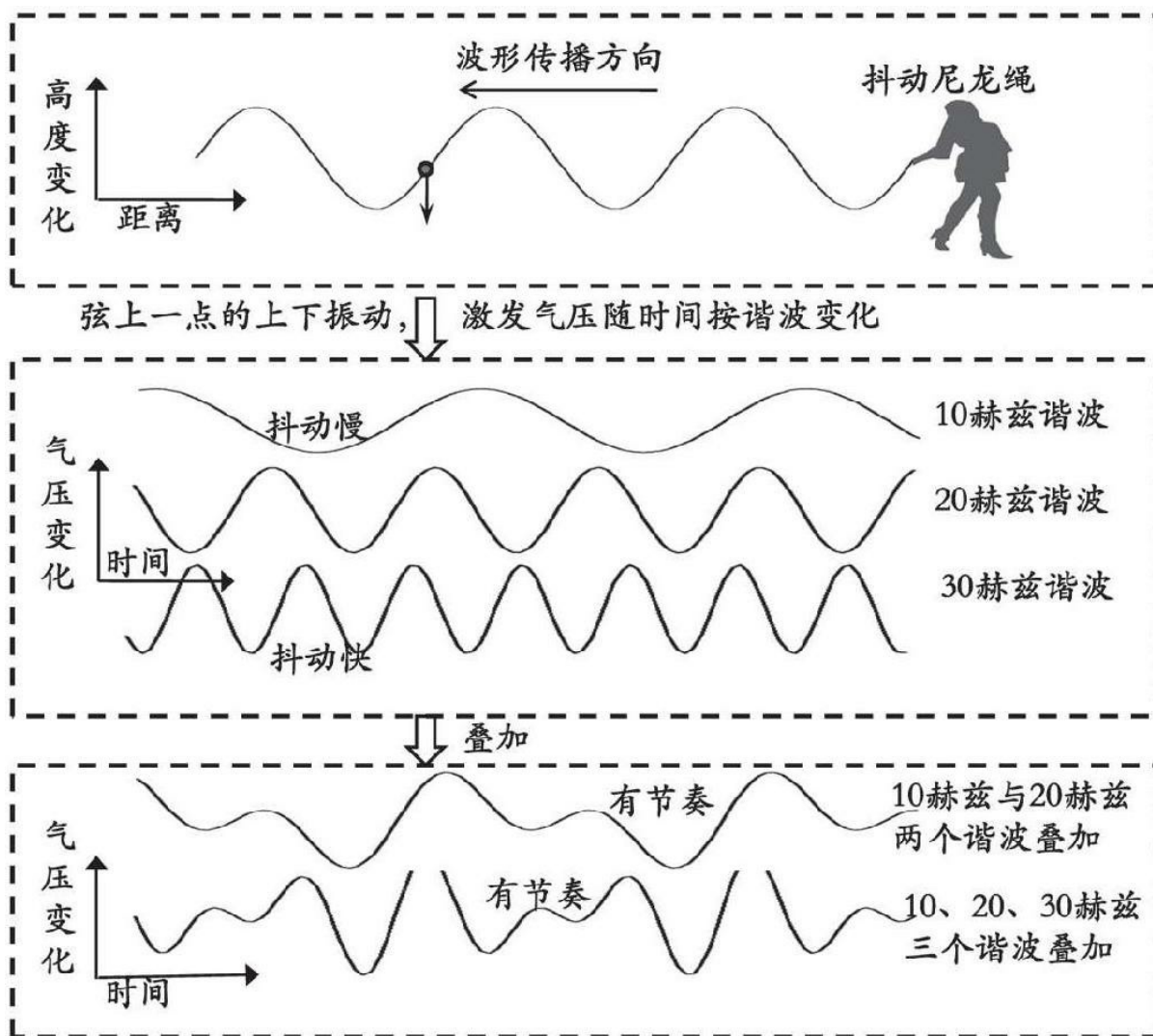


图3.23 具有小倍数比例的谐波叠加出有节奏的气压变化

弦的振动引起气压的变化，这种气压随时间的变化与弦振动时的相对于空间的波浪形态很接近。于是，声音的频率就由弦振动的频率决定了。

琴弦的长度是固定的，按理一根弦在同一时间段只能产生一个基

频。为了使一根弦在不同时间段产生更多的基频，可用手指等压迫琴弦的某位置，使这个位置无法振动。于是可振动的弦的长度降低了，基频就增高了。改变压迫的位置，可形成不同的可振动长度，产生不同的基频，从而引起具有不同基础频率的声音。

按某种规范设计的乐器，在手指调节范围内，能产生的频率会覆盖一个范围。比如说，一根弦自由振动产生200赫兹的频率。用手压迫弦的中点，可振动的长度就缩短了一半，频率就可增加1倍，为400赫兹；压在1/4的位置，就是300赫兹；等等。理想情况下，通过改变压迫的位置，使可振动的长度越来越短，频率会越来越高。但也不能无限缩短，因为那时梅森定理失效了。因此，乐器可使用的频率有一个范围，不同的乐器的频率范围不一样。同一种乐器因设计差异，频率范围也可能不一样（因此不同来源的频率数据往往有差异）。

有的乐器不止一根弦，例如小提琴有4根，钢琴有88根。钢琴的88根弦产生的频率不一样，通过选择不同按键来振动对应的弦，产生不同的频率。

小号之类的管乐，是靠空气在气腔中的振动来产生声音。声波在管道中来回反射，也在一排小口上反射。不同位置的小口反射的声波对应声波一个来回的长度不一样，声波的速度除以这个来回长度得到的频率就不一样。手指封闭不同的小口，就会改变频率，产生不同频率的声音。

注：不同来源的资料，对各种声音频率的上限取法不一样

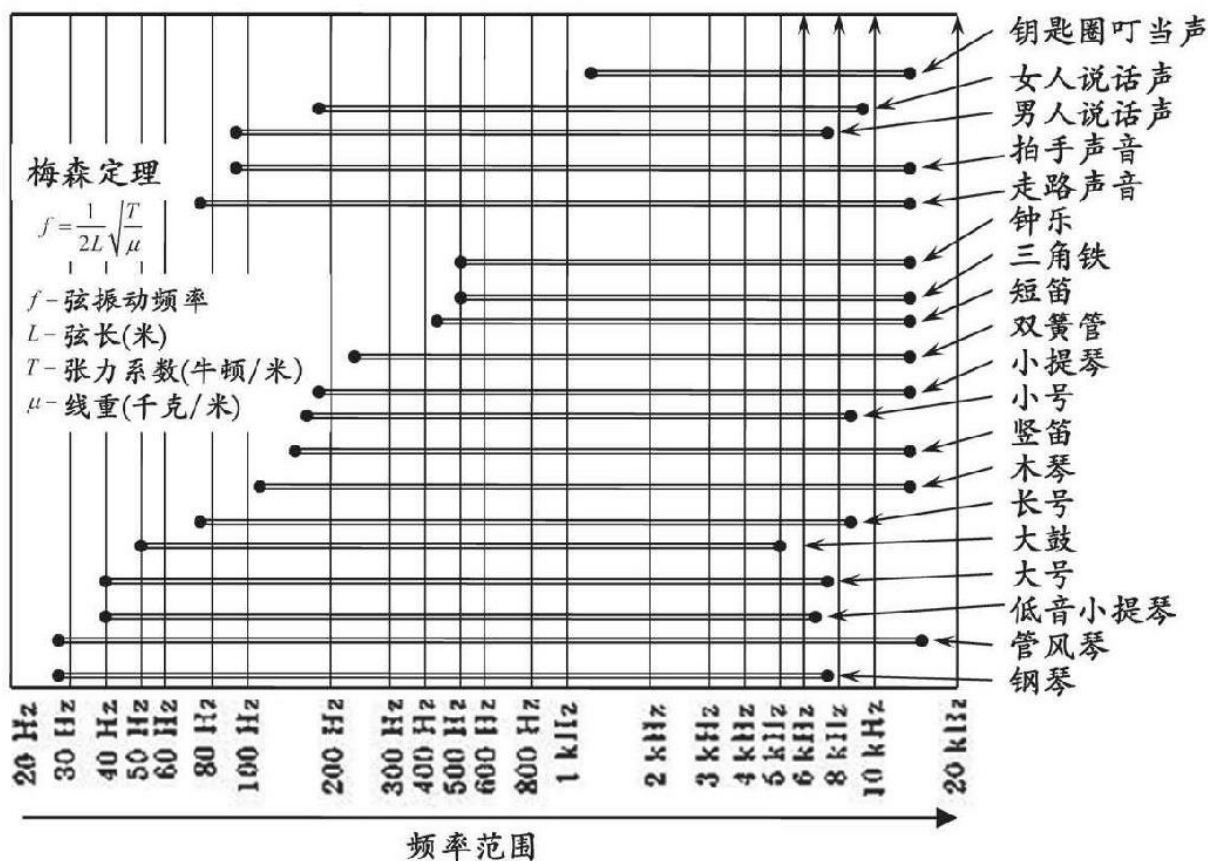


图3.24 不同乐器或声音的频率范围

音阶与频率

无论是唱歌还是玩乐器，我们至少需要熟悉多（do）、来（re）、米（mi）、发（fa）、索（sol）、拉（la）、西（si）。其实，简单而言，多、来、米、发、索、拉、西（简谱分别用1、2、3、4、5、6、7表示）就是七个频率之间相差不大的谐波振动的声音。但这七个频率的选取不是随意的，而是经过长期经验积累以及心理声学对人耳听觉功能研究的结果。

比如说，C大调的七个主音阶“多、来、米、发、索、拉、西”对应

的频率分别为261.63赫兹（记为C4）、293.6赫兹（记为D4）、329.66赫兹（记为E4）、349.23赫兹（记为F4）、392赫兹（记为G4）、440赫兹（记为A4）和493.88赫兹（记为B4）。其中，440赫兹也称为国际标准音。这里，C、D、E、F、G、A、B七个字母是用来给这些频率的声音命名的（因此，C、D、E、F、G、A、B也称为音名），这些字母接的数字4另有含义。如果接5，那么频率就比接4的翻倍。如果接3，就是缩小了1倍。也就是说，这些数字，被递进1，频率就翻倍。

我们人耳有一个奇妙的功能。你把刚才这组C大调的频率翻倍，分别得到523.25赫兹（记为C5）、587.33赫兹（记为D5）、659.26赫兹（记为E5）、698.46赫兹（记为F5）、783.99赫兹（记为G5）、880赫兹（记为A5）和987.77赫兹（记为B5），听起来还是有“多、来、米、发、索、拉、西”的感觉，只是与前一组相比，听起来音调高了一些。这组频率，音名后面接了5，表示与前面那组接4的相比，频率翻倍了。同理，将频率缩小一半，得到的一组频率就记为C3、D3、E3、F3、G3、A3和B3。总之，通过类似翻倍或缩小一半，就可以得到不同的音的频率。比如说，C1就是32.7赫兹，B9就是15804赫兹。从C1到B9，覆盖的频率范围足够大部分乐器的音乐所需要的频率范围了。

并不一定要从C开始，才能得到“多、来、米、发、索、拉、西”的感觉，从D开始也能得到，如按顺序弹奏D4、E4、F4、G4、A5、B5、C5，也有“多、来、米、发、索、拉、西”的感觉。把C作为“多”，就是C大调，如果以D作为“多”开始，就是D大调。类似，有F大调，G大调。

频率按上述方式取离散值，是基于音乐普及、标准化以及心理声学研究而逐渐形成的规范。不同国家的音乐甚至不同乐器也有一些其他规范。这里只讨论了单个频率的选取问题，好听的声音还与频率叠加有

关。在讨论叠加之前，下面先介绍一下与频率比值相关的知识。

比如C4-D4-E4-F4-G4-A4-B4中CDEFGAB代表的频率，不是随意分割的，而是按照某种规则。上面列举的频率，是标准的八度分区对应的频率（即一组音阶，如C4-D4-E4-F4-G4-A4-B4，称为有八度，或者说，下一组与这一组差了八度）。这个八度的“八”字的含义很容易引起混淆，对频率而言，并不是说等分了8等份。实际上，前面列举的频率，恰恰是先按等比关系分成12等份。比如说，把C4（261.63赫兹）到C5（523.25赫兹）之间的频率，按等比关系分成12等份。这12等份得到的频率，相邻两个的比值是2开12次方，即比值大致为1.05946。比如，F4是349.23赫兹，E4是329.63赫兹，两者的比值是1.05946。可是，D4的频率是293.66赫兹，C4的频率是261.63赫兹，两者的频率比不是1.05946，而是两个1.05946相乘得到的1.1224。实际上，从C4到C5中，D4取12个等份中第2个频率，E4取第4个频率，F4取第5个频率，G4取第7个频率，A4取第9个频率，B4取第11个频率。相比之下，F4比E4的频率比值，是其他两个相邻的音阶的比值的一半。于是，称E4与F4之间差了半个音，而C4与D4，D4与E4，F4与G4，G4与A4以及A4与B4之间差一个音。另外从B4到C5，频率比是1.05946，也是差半个音。这样做显得很麻烦，但有其道理，这里不再赘述。

一个音阶的频率翻倍后，听起来像同一个音（虽然高了些），是人耳认知功能的奇妙特点。这种八度等价功能，在恒河猴等哺乳动物中也存在。也许是耳朵鼓膜在声音压力变化作用下，除了按接收的频率产生振动外，还会激发高次谐波（即翻倍的频率），因此与接收到翻倍的频率时产生的听觉效果是一样的。

好听的声音 频率的叠加（图3.23）

频率的大小决定了对应这个频率的声音的音高（音调），气压变化的强度即振幅决定了这个音的音量（响度）。频率越大，音调越高。一组“多、来、米、发、索、拉、西”，按顺序音调越来越高，就是因为频率越来越高。

单一的一个频率形成不了音乐。音乐是一段段不同频率的音通过衔接和叠加等方式得到的。弹奏琴弦时，通过手指的移动，使每个音（频率）有一个起止时间（即在时间上有一段长度）。换一下位置，又得到另外一段音。这样在时间上按某种顺序衔接不同频率（音），就能产生一种音乐的感觉。除此之外，为了更动听，同一时刻还可以叠加不同的频率。例如，演奏钢琴时，几个手指同时按不同的键，就有几个不同的频率的叠加。也可以在同一时刻叠加不同乐器的频率（音）。如何保证每一段音（单一频率）的质量（音质），如何按乐谱柔和地衔接不同频率的音或同时叠加不同频率的音以生成流畅的节奏感，是乐手需要千锤百炼的。

持续按钢琴的一个键或者将一根手指按在琴弦的固定位置不动，那么产生的就是单一频率的声音，声波的气压随时间的变化就是一个波形谐波。通过同时按多个键，使得多个频率的谐波叠加在一起，持续一段时间，形成的声波的振幅（声音的强度）就有节奏的起伏。不同时刻改变频率又产生音调的变化。

具有倍数关系的两个或多个频率叠加，或者频率之比可以约化成小整数比（如 $3/5$ ， $2/3$ ）的两个频率的叠加，产生的声音往往更动听。琴弦按梅森定理产生的谐音是基音，有的琴弦在基音的基础上，还会产生频率翻倍的音（称为泛音）。基音和泛音叠加在一起，就是具有小倍数关系的频率的叠加。

不好听的声音（图3.25）

有几种声音不好听，至少对大多数人是这样。首先是单一的振动持续发出的声音，即只有一个频率的声音持续，就不怎么好听。至少，长时间重复单一频率，不好听。因此说话时，尽量避免把一个字的发音拖得很久（除非你振动一个字时本身就包含不同频率）。

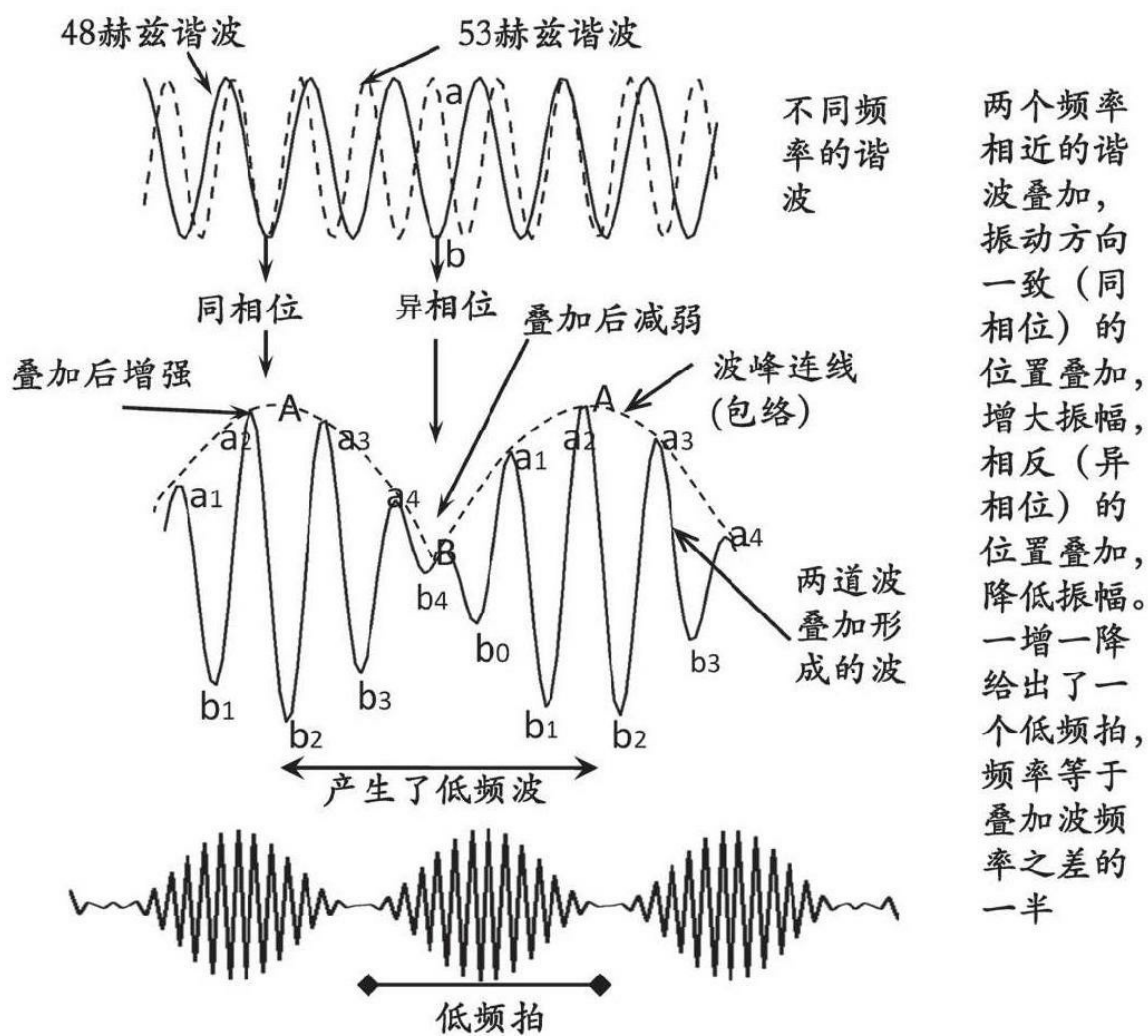
举个单一频率的例子。烦人的蚊子虽然极小，但夜深人静困倦欲睡时，你能听到它的嗡嗡声。晚唐诗人皮日休这样描述：“隐隐聚若雷，嚙肤不知足。皇天若不平，微物教食肉。贫士无绛纱，忍苦卧茅屋”。蚊子翅膀扑动频率高达500赫兹（每秒500次）以上，引起的声波的频率也在500赫兹以上。因此，我们看不清翅膀，却听得到翅膀发出的声音。蚊子周期性地振翅，只发出了单一频率的声音，所以听起来不舒服。

高频的噪声是各种各样的频率的声音毫无节奏地叠加在一起，或者一些频率非常高声音又非常强，听起来也不舒服。

两个频率较高且频率接近的谐波，叠加后，按谐波的数学形式，会出现一个新的低频声音，这个低频声音的频率等于两个频率之差的一半。叠加在高频上，会听到嗡嗡作响的拍打声，比较难听。因此，乐器的不同音阶的频率需要错开足够的值。

振动的频率以及声音频率的分散与叠加，构成了动听世界中的旋律。虽然如此，即使同一个人说的话，或者同一段音乐，有的人觉得好听，有的人就觉得不好听。因此，好听与否，是相对的。这可能来源于不同人的听觉系统之间的差异，对声音中的好听部分和不好听部分敏感程度不一样。除此之外，如果需要从声音中理解一些信息，那么说话就

需要一定的技巧。



两个频率相近的谐波叠加，振动方向一致（同相位）的位置叠加，增大振幅，相反（异相位）的位置叠加，降低振幅。一增一降给出了一个低频拍，频率等于叠加波频率之差的一半

图3.25 两个频率相近的谐波叠加

4. 声源的信息 口语与书面语言

文如其人，是指从文章中可以看出作者的一些个人信息，包括思想、知识水平和一些性格特点。同理，从言谈和声音中也可以了解一些信息，比如说，发出声音的声带应与身高成某种正比关系，而声音的频率与声带长度应该近似成反比，于是，从音调的高低可以判断身高。说

话之中，声音的高低、吐字的快慢和语言的节奏，也如同音乐，掌握得好可能更容易打动听众。演说家声音的音量、音质和节奏给人以响亮、动听和和谐的感觉，高语速是其主要特征。以听清、听懂和消化为目标的教學语言则需要适当结合口语和书面语言。即使是书面语言，也需要许多技巧才能把信息传递清楚。

声音的信息 不同身高的人的音调高低

我们闭上眼睛，可以从声音中识别一些信息。例如，开着小汽车进入隧道后，我们发现噪声加大了。这是因为，汽车激发的声波，打在隧道壁上，会反射。反射回来的声波与原有声波叠加，就放大了噪声（也不排除出现叠加后相互减弱的情况）。

人类说话的声音靠肺、喉部的声带和发音器官共同作用产生。肺部提供足够的空气和气压来振动声带。声带振动空气，发出可听见的脉冲压力波，这就是原始声音。原始声音被口腔和鼻腔这样的共鸣腔放大与调和，再经过舌、颚、齿和唇的配合与微调，最后协调地发出各种音调，形成各种各样的声音。能听到的声音的形成当然比这还要复杂，其中涉及发声声波、不发声声波和爆音的共同作用。

声带发出的声音的频率大约在20赫兹到500赫兹之间。通过口腔其他部位的影响，人说话的声音的频率在50赫兹以上，音量主要集中在300赫兹到3000赫兹之间。人耳能听到的声音的频率在20赫兹到2万赫兹之间，并且对300赫兹至1万赫兹之间的声音最敏感，尤其在300赫兹到3400赫兹之间的声音最清晰和最容易被识别。我们的电话系统采用的频率就在300赫兹到3400赫兹之间。

既然声音的发出源于身体重要器官的协调动作且重要器官的运动能力必然与身体的一些特征有关，那么一个人的声音就有可能透露一些身体的隐私。据生活科学网报道，声音告诉你说话者的性取向、身高、年龄、假声和魅力程度等五件事。

比如，身体越高的人，声音可能越低沉。身材娇小的女性音调更高。原来，如同梅森定理所揭示的那样，人体的声音频率也与声带的长度有一定的反比关系，身体越高的人，声带应该越长，因此频率更低，声音更有可能低沉些。

中医凭望、闻、问、切可以看病，按中医理论，凭声音也可以近似对健康状态进行某种判断。中医将声音概括为呼、笑、歌、哭、呻等五声，认为肝与呼相关、心与笑相关、脾与歌相关、肺与哭相关、肾与呻相关。医术高明的医生也许能从人发出的声音，听出患者的哪一个脏器出现了问题。每一个器官的健康状态也许会对这几类声音的频率、振幅和节奏有影响，也许高明的中医能听出来。

语言的技巧（图3.26）

这里指的技巧，不是巧取豪夺的技巧，不是引人上钩的技巧，不是哗众取宠的技巧，而是准确传递信息让人一目了然完全理解你的意思的技巧。

虽然我们能听到20赫兹到2万赫兹之间的声音，但如果吐字的频率太快，与相邻的字的声音连在一起或者重叠，就识别不出来了。说的目的不是单纯为了把话说完，而是为了让人听。语速太慢，听者不耐烦甚至听不懂；太快，则听不清、听不懂或记不住。因此，必然有最佳语

速。可是，在不少场合，人们说话的目的似乎并不完全是让别人听懂。有一部畅销书，《为何男人不倾听，女人不能看懂地图》（Why Men Don't Listen and Women Can't Read Maps）就提到说与听之间的矛盾。

平均而言，一天之中女人说话更多。在2004年美国有线电视新闻网的访问节目上，艾伦皮斯曾指出：（英语国家的）女人一天能说2万个到2.4万个单词，男人是7000个到1万个。假如一天有10个小时可以说话，1万个单词意味着一小时1000个，平均每3秒左右一个单词。



图3.26 教学语速三种类型

事实上人们说话的时间比较集中。比如说，演说家以打动听众情绪为主，一份15分钟的演说稿一般是3000字，也就是平均每分钟200个字左右。

教学对语速的要求非常高，因为学生听课涉及听讲、听懂和（选择

性)记忆这几个基本环节。听讲首先要求能听清,记忆则需要时间,因此语速不能太快,每分钟可能只能是100字左右。听懂要求意思表达准确,因此课堂语言不能完全口语化,要求结合书面语言和口头语言。但也不能单调地采用一成不变的节奏。例如,强调严谨时可偏重书面语言且需要慢速,每分钟90字左右。强调生动时,可以用口头语言,甚至可以适当快点,如每分钟120字左右。

说话太慢或者在两个句子之间停顿太久,容易导致听众失去耐心。太快或者两个句子之间没有适当停顿,要么听不清,要么听不懂,要么来不及记忆。因此,教师讲课时,只有依据其内容的分类,适当按要求调整语速,才可能达到最佳效果。

课堂传授知识、作演讲与学术报告考虑的是能否让别人听懂。命令与指示的传达往往不会采用书面语言,听到的意思往往跑偏。

美国心理学家达勒(T.L.Dahle)有一项企业管理自上而下过程中信息交流失真的实验结果:董事长信息传到副总经理,信息丢失率达37%,传到企业中层管理者丢失率累计达44%,传到一般管理层,累计丢失了60%,传到基层科室与班组,累计丢失了80%,即从董事长到基层班组,原有的意思还剩下20%。

不同方言或不同语言之间翻译,则可能带来更多的误解。你不妨做一个循环实验,在自动翻译系统中首先输入一句(意思稍微复杂一点的)中文,让其自动翻译成英文,接着把翻译得到的英文再拷入,让其再次翻译成中文,如此继续。你会发现,在有些情况下,中文意思越来越远。

意思的偏离除了翻译不准确外,有时来源于中文可能不强调单复数

和一些对数目、时态与对象的限定。

对于科技语言和其他需要确保意思准确的语言，需要尽可能地给出不可或缺的信息，这些信息尽可能涉及完整的时间、准确的顺序或时态、数目（甚至单复数）、量的单位、从属关系、限定条件等。例如，说到一个物体的噪声是多大时，要求指明是距离多远的噪声，说到多少分贝的噪声会有害身体时，要求指明暴露多少时间才有害，否则给出的信息毫无用处。通知中如果有时间信息，一般同时给出年、月、日、周几和钟点。假如你发一个3月9日开会的通知，如果同时给出是周几，那么事业和教育单位的人就能知道那天有没有时间。信息完整且准确，才能保证不会引起误解。

5. 风中的声音 移动物体的声音 追赶声音的感觉

声音会骑在风中传播，声速叠加风速，就是相对于地面的传播速度。难以想象，如果风速比声速大，那么我们听不到下风口的人说话。超声速飞机还真的比声音快，因此它已经飞到了你的头顶上，你还不知道它来了。驶来驶离的火车，其鸣笛的音调由高到低，原来不是它的喇叭在变化，而是你同一时间段能接受的声波的数目不一样了，或者频率不一样了。假设你说出了一句话，你又能追赶上你的声音，那么，你听到前后顺序被倒过来的一句话是什么感觉？一句话倒着念，没有问题，这句话倒着听，能听出什么吗？

风中的声音 移动物体声音的速度

如果有每秒20米的风在吹，那么站在说话者下风口的人，感觉声音

以每秒 $340+20=360$ 米的速度跑来，而上风口的人则感觉声音以每秒 $340-20=320$ 米的速度跑来。如果风速是每秒340米甚至以上呢？当然我们不可能有这么大的风速，否则下风口不知道能听到什么，尤其上风口的人就听不到下游的声音了。

在没有风的情况下，如果你一边跑一边发出声音，那么声音在静止的地面空气中还是以每秒340米的速度传播。这和水漂激发的涟漪是类似的。不管水漂跑多快，每次落水激发的涟漪，在静止的湖面中，都会形成波纹圈圈以水波特有的速度向外扩张。

物体的移动改变声音的频率 多普勒效应（图3.27）

一个物体以某频率发出声音。如果这个物体朝你移动过来，你听到的声音频率变大了，该物体远离你时，你听到的频率变小了。声波的频率如果越高，那么声音越尖细，或者声调越高。于是，驶近驶离的汽车，其发出的声音的音调先提高再降低。

原来，一辆朝你行驶过来的汽车，后一时刻发出的声波是在汽车更靠近你的时候发出的，相邻时刻声波到达你的耳朵需要的时间间隔相比于静止汽车的就变小了，这相当于声波的原有频率被增加了，于是你听到的声音就比汽车静止时声调要高。而当汽车驶离时，正好相反，因此声调就越低。这就是多普勒效应。需要注意的是，多普勒效应并不说明声波的传播速度提高了，而是频率增加或减小了。上面提到的声音在风中传播，传播速度叠加了风速。



图3.27 声音的多普勒效应

追上你的声音，“回来吧”就变成了“吧来回”（图3.28）

说一句话，声音以声速向四面八方传播。在0.1秒后，以声源的位置为中心，以0.1乘以340即34米为半径的球面上任意位置的人就听到了。0.2秒后，这个球面的半径就是68米。如果有10米每秒的风，那么这个球面包括球心就以每秒10米的速度向下游移动。

假设乘坐某种工具追赶你的声音。显然，以跑步的速度是追不上你的声音的。如果乘坐高速列车，车速是每小时500千米，即每秒139米左右，那么还是追不上声音。

假设你的车比音速还快，那么你就能追赶上并超越之前发出的声波，这时你听到的话就倒过来了。每一个字的发音其实并不前后对称，绝大部分字用辅音连上元音，尤其转舌音和鼻音。不可想象，把一个字

的元音放在前面接辅音，听起来是什么样子。一句话倒过来了，并不是把原有的字按颠倒了的顺序念一遍那么简单，每个字的发音都变了。

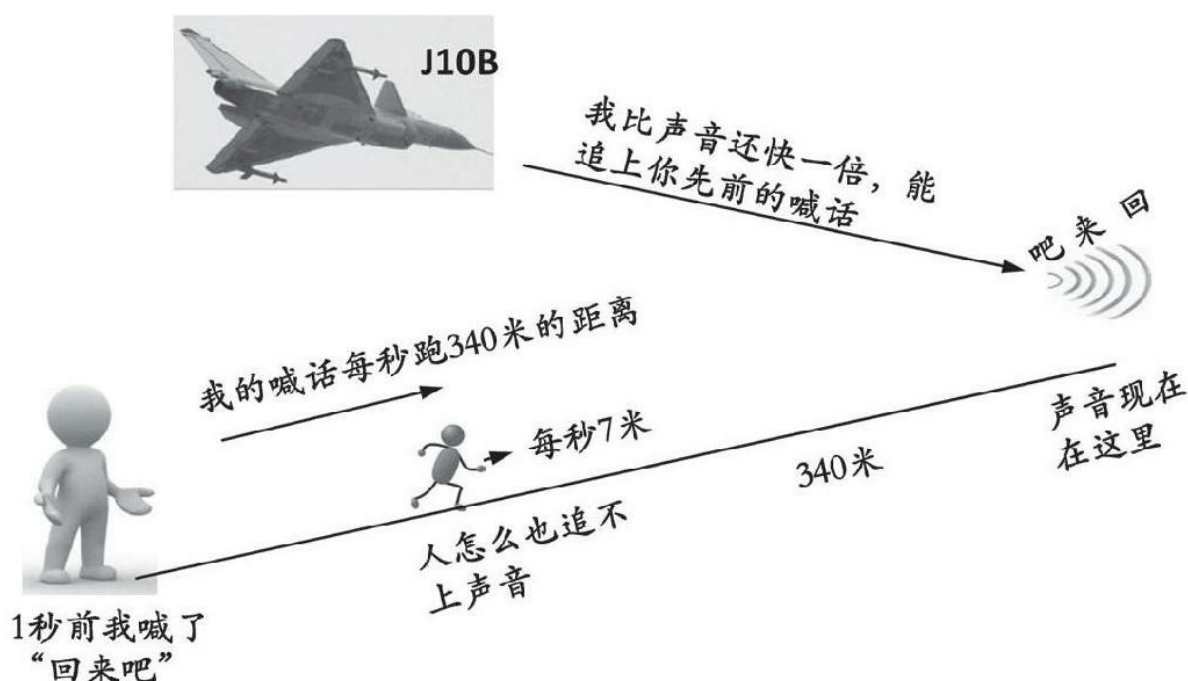


图3.28 追上声音的情况

超声速飞机飞过了，你还没听到声音（图3.29）

一架亚声速客机飞来，远远的还没飞到，在地面的你就能听到它的声音，因为它的声音比它跑得快。超声速飞机比声音跑得快，它飞到头顶了，它的声音还没有到达地面。只有飞过头顶了，过了一段距离，地面的你才能听到声音。

亚声速飞机的声音比飞机跑得快

超声速飞机声音比飞机跑得慢

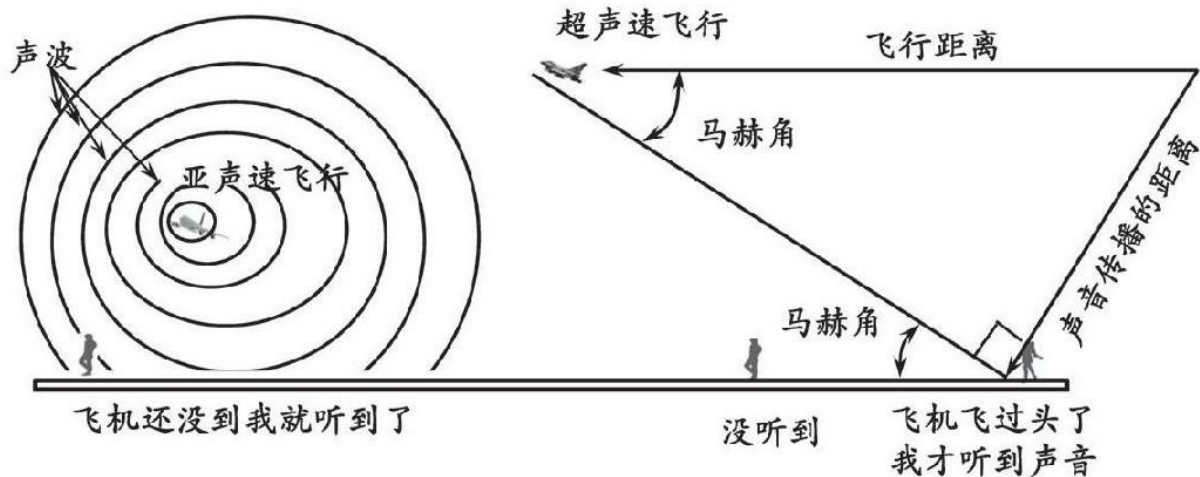


图3.29 在地面听亚声速飞机和超声速飞机声音

考虑在2.4千米高空，飞行马赫数为 $5/3$ （约为1.667）的超声速飞机。约8.8秒多一点点时间内，它能水平飞行5千米，而它发出的声音（按每秒340米的声速，实际比这低一些，因为高空气温比地面低）这段时间能传播3千米的距离。你听到声音那一刻，它已经在离你直线距离为4千米、仰角（等于马赫角）为37度的地方。

这是很粗略的估算。实际上超声速飞机有冲击波，冲击波的速度比大气声速快一些，因为冲击波内的气温比环境温度高，声速比大气声速大，会顶着冲击波以比大气声速快的速度传播。因此，音爆会比普通声音更快到达地面。

3.3 飞行的奥秘

我们一直用了一些不怎么专业的语言暗示飞机如何产生了升力。也不能处处都显得那么不专业吧，于是在这里，我们试图用距离专业更近的语言来说说飞行的奥秘。然而，也不要失望，即使是航空专业人士也会对飞机机翼为何出现升力发生争吵。著名空气动力学家安德森就这样说：“非常奇怪的是，莱特兄弟发明飞机约一百年后，成群的工程师、科学家、驾驶员和其他人会聚集在一起，热情洋溢地辩论飞机机翼是怎样产生升力的；人们会提出各种各样的解释，争论主要集中在哪种解释更基础。”这种争论显然加深了飞机的神秘感。那我们就以探秘者的心态来看看飞行的奥秘。当然，以这种方式看待问题，会在某些本来可以一笔带过的地方，使用很长很慢的描述。我们得从水中开始，再到陆地行走，钻进车里，……，到发动机里面烤一烤转一转，再摸摸高亚声速飞机的腰部，最后随候鸟排队迁徙。怎么要绕那么大弯子啊，不就飞行的秘密吗？

1. 水遇窄道快 人（车）遇窄道慢

这里的主题本来是飞行的秘密，怎么跑到水中与陆地上来了？水的流动，人群的行走，车道上的交通，不需要讲太多道理我们也能知晓其中会发生：遇到窄道，水流变快了，人群变慢了。这谁都懂得的。连这些都懂了，探秘飞行就容易多了。

窄道水流急，狭路人流慢（图3.30）

洗脸池拔掉孔塞放水时，能听到放水的小孔里隆隆的水流声，这表明在小孔那里，水流速度变快了。溪流遇到巨石挡路，那么在石头缝隙之间，水流就加快了。这就是水遇窄道变快现象。宽道中慢慢流动的那么多的水，遇到窄道，要全部通过，当然要跑得更快了。



图3.30 水遇窄道变快，遇宽道变慢

可是，如果公路上车流较为密集，那么遇到道路变窄的地方，车流速度就降下来了，甚至堵得很厉害。人群、兽群也是这种情况。公园、电影院、食堂、教学区的人群，如果得通过狭窄的过道，就会减速并且出现堵塞。

单看这两种情况，都很容易理解，都在我们常识之中。可是，把两种规律相反的现象放在一起，就不好理解了。流过飞机的气流，亚声速时遵循水的规律，超声速时遵循人的规律。原来，人群的行走居然属于某种意义上的超声速流动，只是这个声速不是声音的速度，而是类似于声音那样的人与人之间的相互干扰的传播速度。

车流与水流中的扰动

为何车流遇到窄道会减速，水流遇到窄道会加速？原来，遇到窄道也相当于车流与水流是受到妨碍了，水流遇到扰动会快速将其释放完，而车流遇到的扰动会来不及释放而妨碍车流。

道路或水道突然变窄，就是一种干扰源。现在用我们熟悉的道路上

的车流为例来看看发生什么。试想你在开车，你的前面道路的变窄（或其他原因）会导致你前面的一些车并道、减速或出现其他会引起你作出响应的状况。这对你而言就是扰动。遇到扰动，你有一个从看到扰动和意识到扰动，到手脚作出动作使车作出改变的过程，历经的时间就是总的反应时间或响应时间。

对于普通司机，这个响应时间是1.5秒左右。有的司机可能得3秒甚至更长。赛车手当然比1.5秒短多了。

为了说明问题方便，这里取3秒作为平均反应时间，前后两辆车的平均车距是30米。遇到扰动后，一辆车作出响应，30米之后的那辆车的司机看到这种响应（如变速、并道等），在3秒后完成同样的响应。接着，这种响应进一步往后传。这就是说，扰动以30米除以3秒，等于每秒10米的速度在传播，即扰动传播速度是每秒10米，每小时36千米。

当然，这里假定了响应时间都是3秒，车距都是30米。实际上不同的司机的响应时间和车距可能差异很大。这种简化讨论至少让我们了解到扰动会以某个速度传播。

车速与扰动速度的博弈

于是，就出现了两个速度。一个是大多数人知道的车速。另一个是大多数人不太熟悉的扰动或响应传播速度。当你遇到两个不同类型的速度，你首先要做的，就是比较它们的大小。飞行速度与音速之比称为马赫数。于是，可以把车速除以响应传播速度所定义的比值，称为行车马赫数。对于行人，可以称为行走马赫数。

继续考虑每小时36千米的响应传播速度情况。如果车速都是每小时

18千米，那么行车马赫数就是0.5。如果车速是每小时72千米，那么车速比响应传播速度快一倍，行车马赫数就是2。行车马赫数大于1的叫超波速速度，小于1的叫亚波速速度。

对于其他任意车速或车距，都可以得到类似的行车马赫数。如果是人群行走，也有行人之间的间距和反应时间，两者之比也定义了一个响应传播速度。行走速度除以响应传播速度，就是行走马赫数。

事实上，道路上之所以出现车流和人群，表明大家都着急，都急着往前赶。因此，一般情况下的行车马赫数和行走马赫数都会大于1。遇到扰动，比如说遇到窄道，由于扰动传播速度比移动速度慢，因此就会干扰移动，迫使减速。你前面的人减速了，你还以原来的速度前进，就会撞在一起。人（车）遇窄道慢，就是这个原因。当然我们也不要因此激动不已，庆幸自己都能超“音速”了，“马赫数”都大于1了，要知道，你的这个“音速”不是那个声音的速度。

水的流动也会遇到扰动。这个扰动传播速度就是水中声波传播速度。水中声波的传播与大气中的声波有点类似（与水面波浪的传播速度不是一回事，是水受到挤压和拉伸产生的波动的传播速度）。这个速度大致是每秒1500米，即每小时5400千米，比大气中声音速度快多了。水流速度一般会较低，比如说每秒15米。每秒15米，除以每秒1500米，这样得到的水流马赫数是0.01。于是，有什么扰动试图造成堆积，就会通过压力波以每秒1500米的速度极快地释放到别的地方。因此，水的流动就不必顾忌有什么扰动挡路，遇到窄道就会遵循体积流量不变（即单位时间通过水道的总水量），选择快速通过。

因此，人与车的移动速度一般比相互之间的干扰的传播速度快，遇到扰动后，扰动就添堵，迫使人与车为了避让就减速，一下子在干扰区

域堆积，从而会拥挤在一起，速度就变慢了。水流的速度比干扰速度即压力波慢许多，有点什么干扰，比如说水道变窄带来的干扰，这种干扰的影响很快就传播到足够远的地方，让水流及时作出调整，不会出现堆积现象。

就是说，对于人车流动，扰动比你的速度慢，就给你添堵，嫉妒你。对于水的流动，扰动比你速度快多了，遇到窄道就出来帮忙，把对你的妨碍快速扩散到别的地方。

2. 风中转掌知升力 优美的流线与和谐的声波

不要羡慕鸟的翅膀和飞机的机翼能产生升力，因为我们的手掌就是机翼，也能产生升力。别害怕，不是要你把手掌插在飞机上在高空去飞（那多危险），而是让你舒舒服服坐在车里，用手掌去感知有升力的乐趣。升力当然来自于气压差，气流速度的差异提供了这种气压差。气流速度的变化和气压都来自于分子热运动，它们是如何把一部分当作吹风一样的气流，同时把另一部分当作气压？气流和气压是在友好相处，还是相互为敌？如果互为因果关系，谁是母鸡谁是鸡蛋？

外探手掌知升力（图3.31）

设想你坐在行驶的小汽车右侧的座位上。在确保安全的情况下，将右手伸出右侧窗外。首先将手掌放平，那么你顶多会感觉到风给手掌向后推的力，且只有车速足够快，才有明显的这种感觉。

现在你将右手手掌顺时针转一个小的角度，即让手掌有点迎角。如果车速较小，或者迎角较小，那么不会有奇迹发生。如果你手掌心正对

着前面，即迎角等于90度，那么就没有升力，只有阻力了。此时，手掌就会被推向后面。

现在让车速快点儿，比如说每小时70千米。将五指并拢，手掌尽量展平，接着将手掌迎角慢慢增大。会出现一个迎角，可能在30度左右，你的手突然会被举起来。

就是说，手掌获得了向上的升力，至于迎角多大会把手掌抬起来，还与手掌的大小、车速以及手型有关。当然，如果在车上做这种体验，一定要注意安全。即使外部安全，也要慢慢增大迎角，别因为你手掌的形状太利于产生升力，以致突然产生很大升力，抬起速度太快而磕着车窗框。总之，宁可不去体验升力滋味，也要保证安全。

你可以调节手型和迎角，使升力足够大，再感觉一下手心、手背以及正对着前方的拇指会感受到什么。有些地方气压高了（比如说拇指靠掌心的部位）、有些地方气压低了（手背拱起的地方）。尤其是作为前缘的拇指上侧，那里负压很大。

手心手背并不是对称的，手心窝进去，手背向外凸。这种弧度也称为弯度。除了迎角会产生升力外，弯度也贡献升力。你可以改变手型凸起的程度，看看升力是否变化。

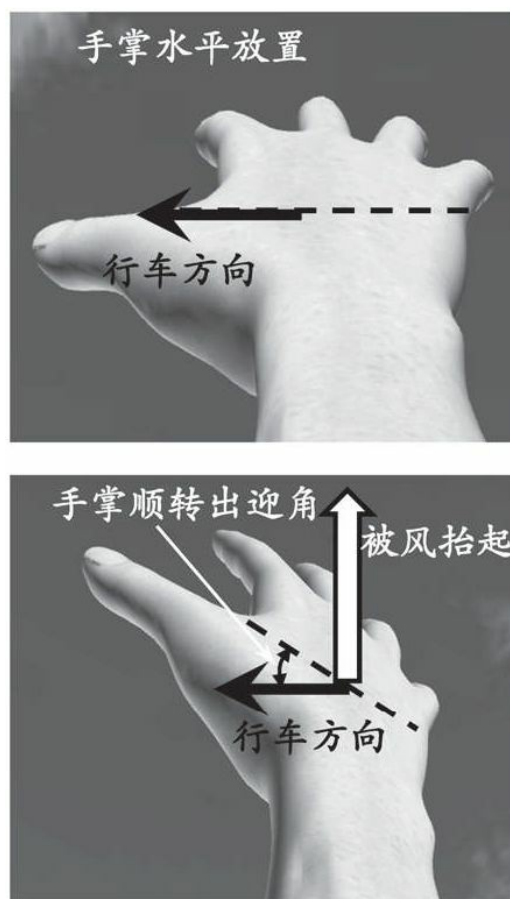


图3.31 手在风中产生升力

弯度分担迎角的职责（图3.32）

人的手掌往往是掌心凹进去一点点，手背有点凸。如果细看鸟的翅膀，就会发现翅膀向上带有一定的弧度，即向上凸起。这就是弯度。鸟扑动翅膀、人的手抓握物体，从肌肉进化的角度不难理解会形成这样的向一侧凸起的弯度。不仅如此，鸟翅膀的弯度还能分担一部分迎角的作用，也能产生一些升力。

为了产生足够的升力，迎角必须足够大。迎角独自承担升力任务，太累了，弯度过来帮帮忙，一起产生足够的升力，这样就不会出现迎角太累而趴下（升力下掉）的情况。

至少可以这样理解，气流绕凸起部分流过时，走了弧线，离心力降低气压，于是翅膀凸的一侧也会产生负压即吸力，从而贡献升力。

迎角效应和弯度效应都有利于产生升力。两者都利用，就有可能取长补短，达到最佳效果。没有弯度单靠很大的迎角效应产生升力，会迫使气流急急地绕前缘拐弯，容易产生拐弯涡，或者因拐弯离心力太大，导致前缘产生极高的负压峰值。负压集中在一个点当然不舒服（就像你睡在一颗钉子上）。让迎角适当小一些的弯度则可以让拐弯半径大多了，拐弯涡不容易形成。弯弧的上表面导致的吸力（负压）能均摊在更宽的范围，在机翼上的气压分布更均匀和更柔和。因此，弯度不是在抢迎角升力的饭碗，而是与迎角效应合作，使得在产生升力的同时，构成升力的吸力能柔和地分布在更宽的表面上，同时避免迎角单独挑重担时产生拐弯涡的情况出现。

机翼和翅膀也像手掌一样向上凸起（称为有弯度），迫使气流走弧线产生离心力，降低凸起部位的气压，产生贡献升力的吸力

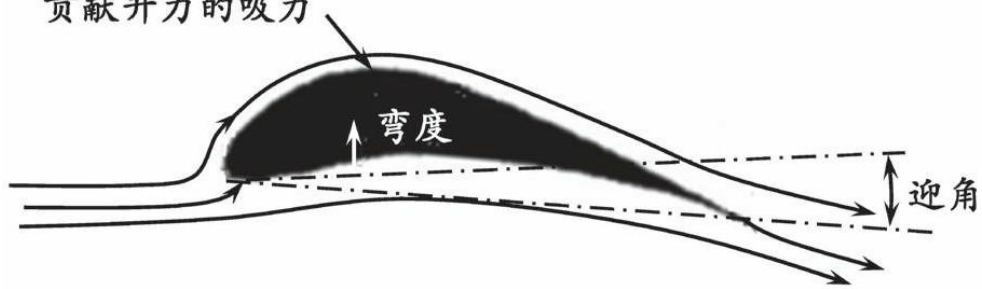


图3.32 弯度也贡献升力

烟雾萦绕 水流湍急（图3.33）

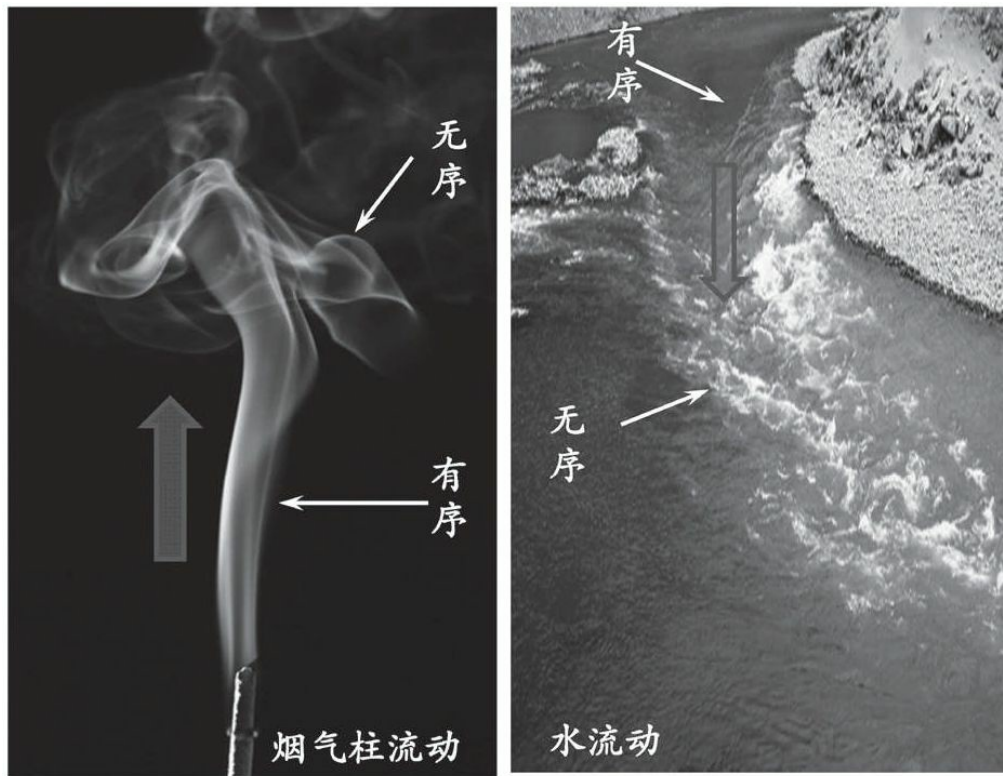


图3.33 有序流动（层流）和无序流动（湍流）

香烟燃烧加热了周围的空气，使它们变轻，因此在浮力作用下上

升。上升的区域正好带着白色或黑色的气体（二氧化碳或者杂质），因此我们看得清流动的样子。

水在河道中流淌，如果遇到巨石挡路，那么水面波浪的变化让我们看得清流动的样子。形容水流有滔滔江水奔腾不息、一泻千里、波光粼粼、碧波荡漾、流水潺潺等词句，可见文人之笔能惟妙惟肖地再现水的流动形态。

香烟柱的初始段显得非常有序，顶多轻微摆动。到了一定的高度，就乱了。水流也存在有序的和无序的区域。这种有序的部分，专业上称为层流；无序的部分，专业上称为湍流。烟雾萦绕，水流湍急，可见湍流已经在文学语言中出现了。

一小团烟雾和一团团水，在湍流中忽上忽下、忽左忽右、忽前忽后，速度忽大忽小。这有点像分子热运动。我们在讲高尔夫球时说过，好好的有序层流，为何会发脾气一样变成湍流。原来，维持层流，每团流体活动范围太小，不舒服，变成湍流，每一团流体就可以轻易跑到更宽的范围。不被囚禁在窄小的空间憋死，也是水的权力，空气的权力嘛。你想想人会怎样，就能理解水和空气会怎样。话虽这么说，往下讲机翼时，我们会知道，湍流之所以产生，也是因为一些因素在内斗。在一起吵起来了，还不如跳到别的地方。惹不起还躲不起？

有了上面的直观感受，就容易想象空气的流动了。因为道理是一样的，只是空气流动看不见罢了。但是，我们还得用合适的眼光去看气流。

假如我们用微生物的眼光去看空气流动，那就复杂了。微生物目光短浅，它们可能看到的就是一个一个的做热运动的空气分子，在那里跳来

跳去。假如我们带着超倍的放大镜，去看照片、画、脸部，会看到什么？会看到一个个的坑。我们不会这样去欣赏照片，看人表情。

我们看得清眼睛、鼻子、嘴巴就够了，当然看得清睫毛更好。我们最好不要看得清细胞。当然，我们仔细看，还是看得清脸部和手掌上的纹理，正是这些纹理让我们有了平面感和空间感。

我们就以这种看清纹理而不去看清细胞的眼光看气流就够了。气流中可能有一定大小的旋涡，那就当它是脸部的眼睛就行了。就这么简单。

握紧拳头时，脸部微笑时，皮肤上的纹理就会有特定的形态。流过机翼的气流，类似地也有一定的形态。脸部因做表情在动时，这里绷紧那里舒展了，感觉这里在拉扯那里在松弛。气流也是，这里慢了那里快了，气压这里高了那里低了。气流速度和气压，均代表了空气分子的某种能量，从而它们的变化得相互顾及，不能擅自行动。这样才能让流线优美，让声波和谐。

流线的优美与声波的和谐（图3.34）

在气流中，组成小团的分子们除了和微小的小气团一起旅行，自身还在做热运动。如果随小气团一起走得快点，自身的热运动速度就降低一些，因为不会无缘无故跑得更快了。你奔跑的速度越快，那么用于挑担的气力就越小。空气也是人，岂能一边跑一边还有那么高的气压。

于是，在速度较低的情况下，这种流速与气压的变化趋势相反，使流速代表的动能与气压之和是常数（这个原理在专业上也称为伯努利定理，是伯努利最早发现了这个原理）。

如果气团的速度较高，与音速相比不是特别小了，那么动能和气压之和就不是常数，但还是会满足速度越快气压越小的趋势。

流速加快与气压变小是一种相互依赖的关系，不是前因后果关系，而是并列关系。在一个空间区域，流速不一样气压就不一样，因为要满足二者之间的协同关系。那些流速大的地方，气压小，反之亦然。从一点出发的气流小团，可能走直线，也可能走弧线，甚至走圆周线（如漩涡中的小气团）。从相邻位置出发的两个气流小团，走的路径不一样。两个路径可能越来越靠近，或越来越远离。靠近时，里面的气流就得加速了（亚声速时，水遇窄道变快嘛），气压就变小；或者减速了（超声速时，人遇窄道变慢嘛），气压变大了。

这种路径，就是流线（专业上，流线的定义比这复杂点，但对于不随时间变化的流动形态，这种路径就是流线）。说一个物体是流线型，是指其表面正好是流线，即小气团能始终贴着物面走（而不会在某位置甩出去）的外形。

你可以想象，流线该具有优美的空间形态，这里直直的，那里弯弯的，还有的地方是圆圈，就像艺术家的素描画一样。

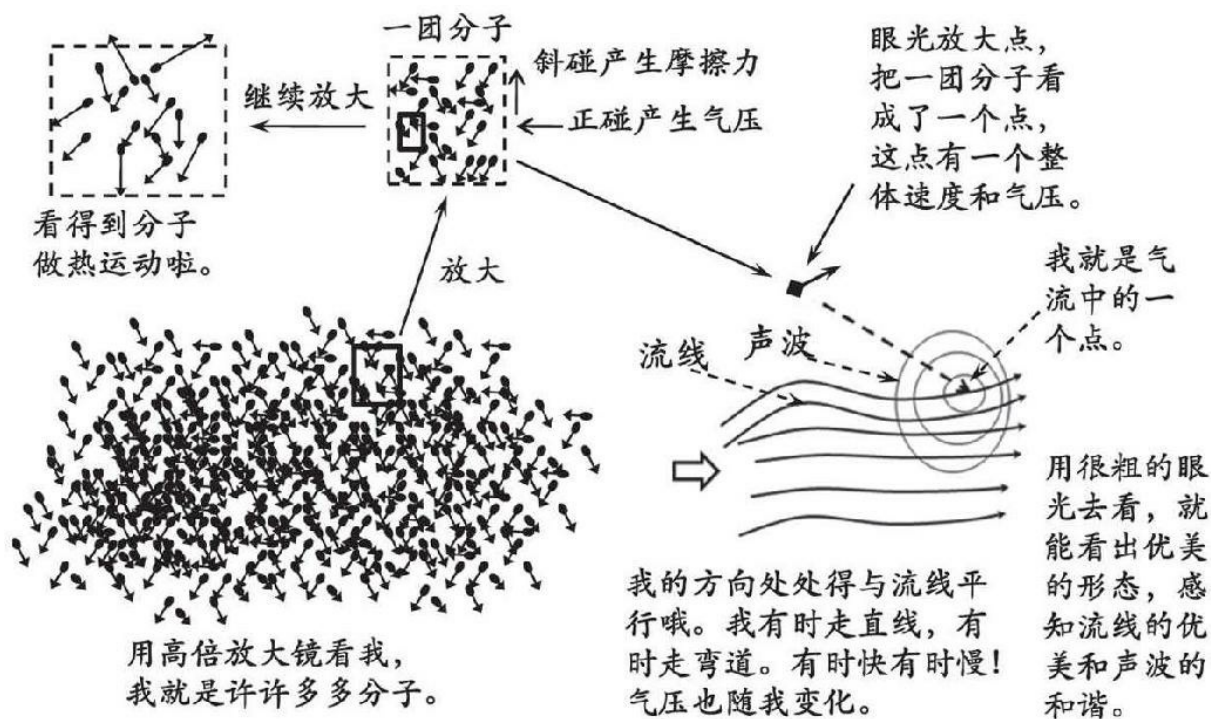


图3.34 不同眼光看气流

由于流速和气压是相互依存的关系，因此流线分布有多优美，气压的变化就有多和谐。飞机要设计成什么样子？让绕飞机的气流中，流线优美、声波和谐。一切都是为了美，好的飞机一定是好看的飞机。

3. 在机翼的气流中旅行

站在飞机上看，气流绕行机翼一定很神秘。那就让我们化作一团气体，跟着气流在机翼周围旅行一圈，体验一下这种旅行的快乐与煎熬吧。不过，我们得有所准备，因为这个旅行有点漫长。你得忍受被禁锢在一处的感觉，忍受与同伴的离别之恨，更得忍受在一起的小伙伴相互推搡摩擦导致的混乱。当然，你更多是在体验过山车的滋味。无论是从下方举了机翼一把还是从上方拉了机翼一把，你在离去之时总算为飞机不掉下来做了点贡献。

绕行机翼的旅行者（图3.35）

我现在是机翼上游的一小团空气，在向机翼走去，开始我的旅行。旅行中的任何时刻，一些空气分子会无情地离开我，但也有一些会很友好地加进来。我如同公司老板，任由自己的员工活蹦乱跳地进进出出，但那些厮守在里面的，还是随我一起向右移动，人们把我走的路径叫迹线（如果不随时间变化而只随空间有变化，迹线就是流线）。

如果我在某处的速度小于声速，我会遵循“水遇窄道快，遇宽道慢”；如果高于声速，也会遵守“人（车）遇窄道慢，遇宽道快”。这是大自然赋予我的法则，我不可能突破。

现在我跑不过声速。我从机翼头部（前缘）正对着的上游在通往机翼的路上，靠近机翼时，我得减速。人也会这样，遇到挡路的就会减速。

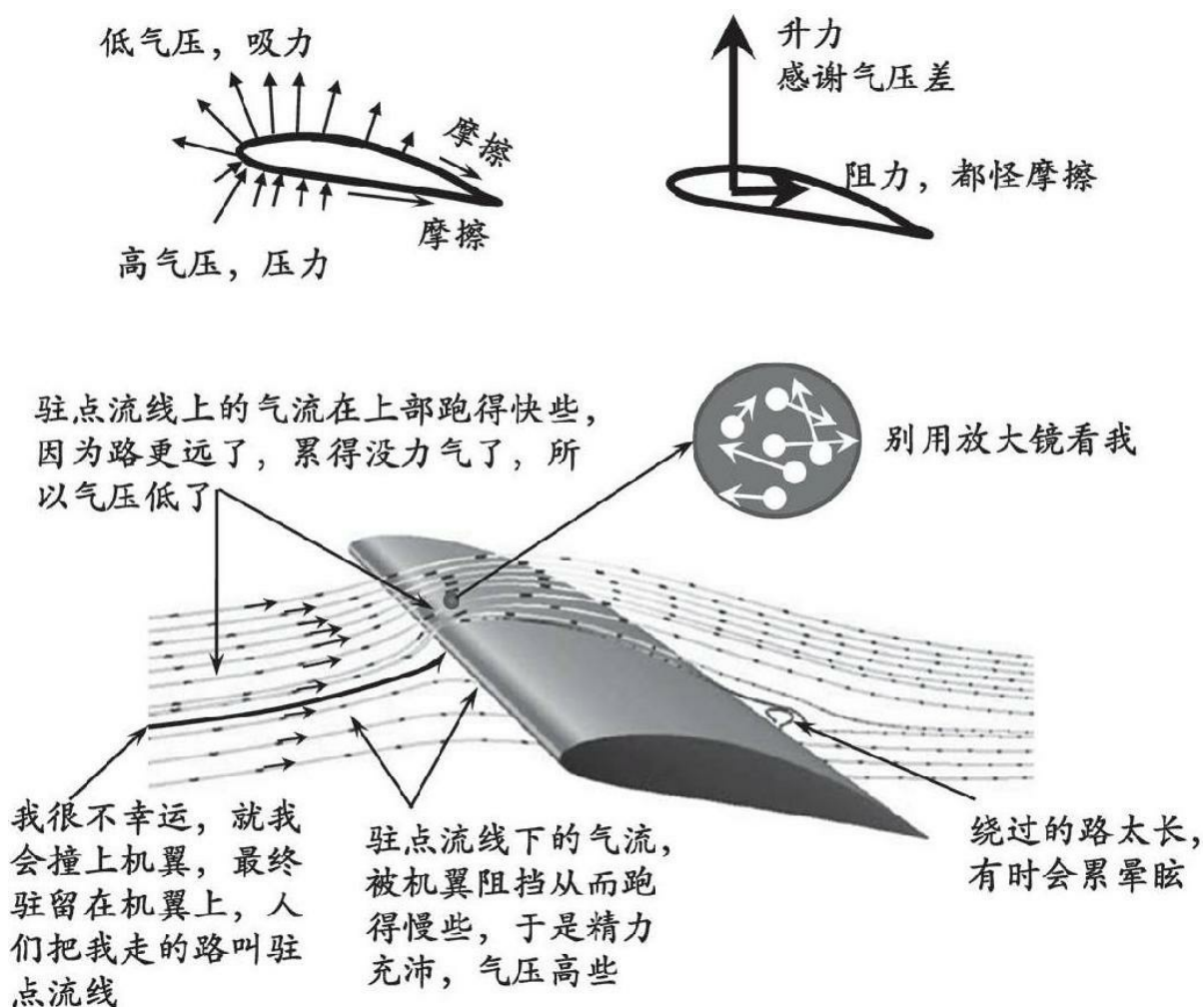


图3.35 跟随机翼一起看气流

我如果不走运, 走的路径使我恰好正面撞上带迎角的机翼, 我就被迫停在撞击点上了。我没法回头, 因为后面的空气会堵上我。人们把我走的路叫驻点流线。我将永远留在机翼上了, 囚禁我的那个点也叫驻点, 我驻留在那个点上, 不能动弹了。没有了速度, 因此气压就高了。

如果我从上游出发时, 位置比驻点流线高一点点, 那么我到了前缘后, 就得绕过前缘, 去机翼上表面继续我的旅程。我绕过前缘时, 也相当于水遇窄道了, 因此我跑得就很快, 快了后我的气压就很低了, 因为也累嘛。比大气的气压低许多了, 因此机翼就受到我的吸力, 看来我的

辛苦没有白费。

距驻点流线更远的在我上方的那些邻居们，没有像我这样受到机翼那么大的阻挡，因此气压接近大气压。于是，气压大的邻居也会压迫我，将我朝机翼方向推挤，帮助我克服了绕前缘时拐弯的离心力。

也可以这样来理解。我拐这么大的弯，受那么大离心力作用，显然只有机翼前缘拖住我，我才能绕过去。这个拖住我的力就是吸力。前缘吸我，我就通过降低我的气压吸机翼前缘。这也是种礼尚往来。

好不容易绕过了前缘，但还得加速才能到达机翼的最高点。从机翼最高点开始，我的道路开始越来越宽了，于是我就减速了，水遇宽道慢嘛！我不那么着急了。于是，我的气压又开始回升，终于可以喘口气了。最后，我到达了尾缘。那里是尖的，便于我与那些从下面来的伙伴们相会。我能见上出发前高度比我低一些但和我一同出发的伙伴们吗？

非常令人失望，一分手就成了永别，和我一同出发的位于驻点流线下面一点点的那些伙伴，会比我晚到尾缘。我都走到尾缘下游去了，它们才到尾缘。出发时就相距那么一点点，这一别就是永别了。我只会见那些比我早出发的了。我的位置就高了一点点，居然享受高速列车的待遇，比我的地位低一点的伙伴，就只能坐慢车了。

可是，也不能用地位高低和速度快慢来贬低我的伙伴。在我下面的那些速度慢的伙伴，精力充沛，气压大，让我羡慕极了。再说，它们不用绕那么大的弯子，不会像我似的累得一点气力都没有。尤其我到了机翼后段，我的气力小得可能会被甩出去。一旦甩出去（气流分离），我就会被卷进很难抽身的旋涡了，高处不胜寒啊！

如果我有幸沿着流线型机翼（那种纤细、薄平并且抬起的角度不大

的机翼)旅行,我就不会拐太大的弯,因此体力消耗再大,也不容易出现晕眩。

我在绕过上表面时,尤其在前缘附近时,由于降低了气压,因此给机翼产生了吸力。那些在机翼下面慢慢旅行的伙伴们精力充沛,提升了气压,给机翼下表面产生了正压。我们虽然分道扬镳,但还是齐心协力,上拉下顶地给机翼提供了升力。

当然,我们在机翼前端威力大一些,那时走的路还不多,体力消耗小,因此升力主要作用在机翼前半段。

贴近机翼的旅行 最接近诺贝尔奖的流动世界 (图3.36)

如果我完全贴着机翼壁面走,我也玩完了。这是因为,机翼会通过摩擦阻止我旅行。除此之外,离机翼更远一些但靠近我的伙伴们,如果想比我走得更快,那么我们也会发生摩擦。于是,邻居伙伴的速度也降下来了。还好,稍微远一点的伙伴,受到的摩擦小多了。于是,如果我们在贴近机翼表面的一薄层内旅行,速度就会降下来。这一层叫摩擦层(专业上叫边界层或附面层)。这个摩擦层是空气动力学之父普朗特教授的伟大发现,据说这个发现及其所涉及的数学模型的建立,差点使他拿了诺贝尔奖。于是人们说,摩擦层的发现是飞机那样的气流问题最接近诺贝尔奖的工作!

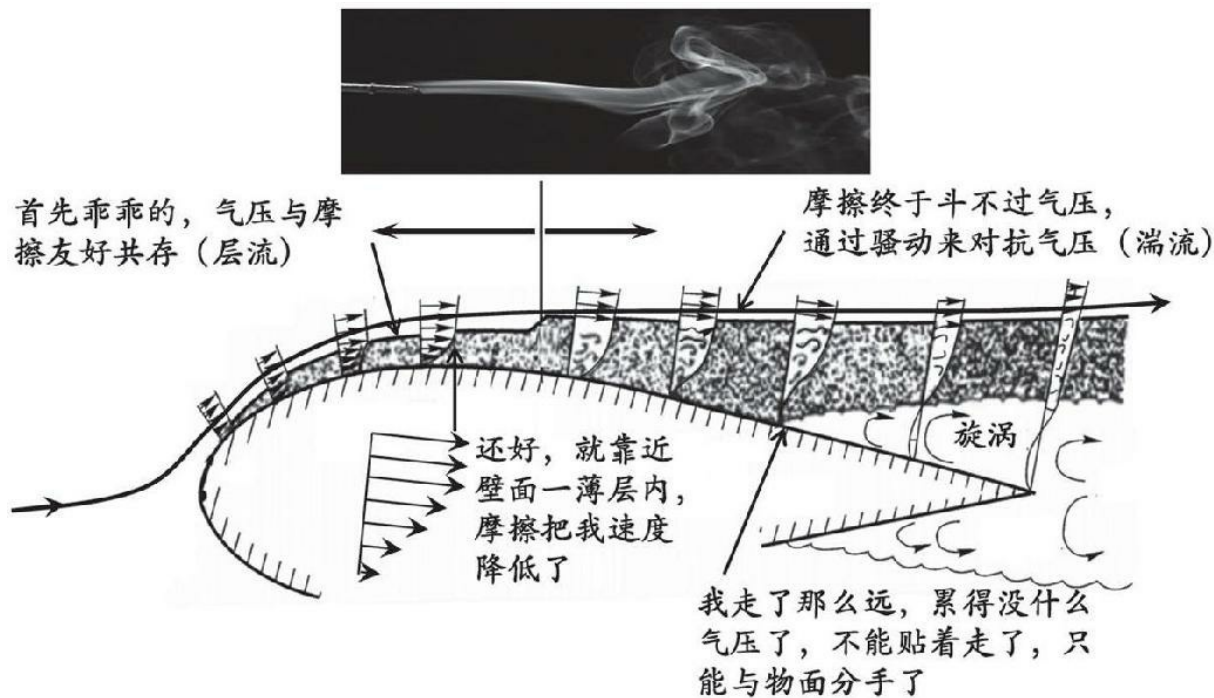


图3.36 边界层：气流与机翼的摩擦在很薄的一层内降低气流速度

我在摩擦层中旅行，不但会受到我的前后邻居用不同的气压来挤压或拉扯，也会受到我的上下邻居的摩擦。这让我够受的。

我们越往下游走，上面越多的伙伴受到我们的摩擦，从而被摩擦的一层越来越厚了。我上下被摩擦，前后被挤压，人越来越累，都站立不稳了，因此，稍微遇到扰动，我可能就趑趄起来，将邻居们绊倒，导致它们也趑趄起来。

我们终于一片混乱，人类把我们叫湍流。那个状态，就像前面介绍的烟雾萦绕，就像水流湍急的地方。湍流是一个奇特现象，我在湍流中，一会儿跑到这里，一会儿跑到那里，杂乱无章，随机变化。有的人因此认为我会忍术。其实，是我和我的邻居们颠三倒四地乱了方寸。我们有多乱？乱得最伟大的物理学家和数学家都无法抓住我们的规律。多少科学家为此毕其一生啊！我又是物理学最后遗留的难题，又是数学无

法解决的难题。不信？那就去看看克雷数学所网站，人们把我作为千禧年七大难题之一，悬赏100万美元，让你挖开我的秘密。

我在发威时，飞机都会颠簸。空姐不时提醒乘客：“飞机遇到一些湍流，请系好安全带”。当然，我们也不是一无是处。如果我们被试图甩出去，我们一个趑趄，就会回到原位。高尔夫球就通过布满凹槽，可以让我们变成湍流来阻止拉扯它的拐弯涡，使高尔夫球飞得更远。我们变乱后，也会偷偷地把外面速度快的气团带近物面，使物面附近气压不会太低，避免拐弯涡的产生，减小拐弯涡带来的阻力。当然，我们这么趑趄，也会加剧与机翼的摩擦。

人类试图用失稳来解释我们为何变成湍流。人类在走路开车时，会受到某种扰动，如风吹草动、虫飞鸟鸣、地面不平、路人干扰以及光照改变等。人类受到一点干扰就会情绪失稳暴跳如雷吗？骑车时遇到一点风吹草动就会歪歪斜斜、跌跌撞撞，甚至摔倒吗？都有可能。那就叫失去稳定了。失去稳定，可能会乱作一团。

原来，我之所以乖乖地旅行，凭的是惯性。我的前后伙伴会推我、挤我，或者拉扯我，我的上下伙伴会摩擦我。开始时，我还能忍受这些骚扰。但到了后来，我的体力不行了，老了，脾气变古怪了，因此我们会骚乱，人类看起来，我们就是湍流了。

进入摩擦层不久，在我前面的邻居提前加速，气压比我低，因此我能顺利往前走，上下邻居对我的摩擦也不会给我造成太大麻烦，我这段旅行可谓顺风顺水。我的伙伴们和我一起，很有秩序地旅行。人类把我们叫层流，即各走各的道，各在各的层。

可是，过了机翼的高点，我前方的伙伴们的气压升高了，于是对我

有所妨碍。我有点走不动了，因此我们更有可能变得混乱。

命运决定于机翼的形状与迎角（图3.37）

如果机翼的前缘太尖且迎角较大，那么我绕过前缘时，就会遇到很大的离心力。我又没有系安全带，因此就会甩出去。甩出去又被上面的伙伴推回来，因为他们没有经历甩出去的痛苦，有的是气力把我推回来。当我被推回到机翼时，我已经到达了机翼中段附近的某位置了。我在甩出去与推回来的过程中，会带动我下方的气流旋转起来，人类把这个旋涡叫前缘涡，因为它靠近前缘。其实就是一种拐弯涡。

这个前缘涡也像生命，一旦产生，就会长大。这是因为，除了我被甩出去，我后面的伙伴绕前缘时，也会被甩出去。大家都带着这个旋涡转，就慢慢长大了。前缘涡长大后，就不会留在前缘了。男儿长大志在远方嘛。长大的前缘涡就会脱离前缘，跑去尾端看看。

脱离原位的拐弯涡，走近尾缘后，会把尾缘上的气流也带着转起来。尾缘下方的气流于是就得绕过尾缘来迎接，绕过尖尾缘也会遇到离心力，生成尾缘拐弯涡。于是，前缘涡就有了双胞胎弟弟。他把弟弟带大，继续他的旅行。尾缘涡弟弟长大后，也会追随哥哥，开始他的旅行。

前缘涡离开前缘后，其他从上游过来的伙伴绕过前缘，会继续滋生新的前缘涡。于是，前缘涡和尾缘涡会成对地产生。到了下游，它们排成两排。上一排顺时针旋转，下一排逆时针旋转，像一个队列。人们把这个队列称为卡门涡街。之所以叫卡门涡街，是因为卡门发现并研究了这种涡街引起的额外阻力。

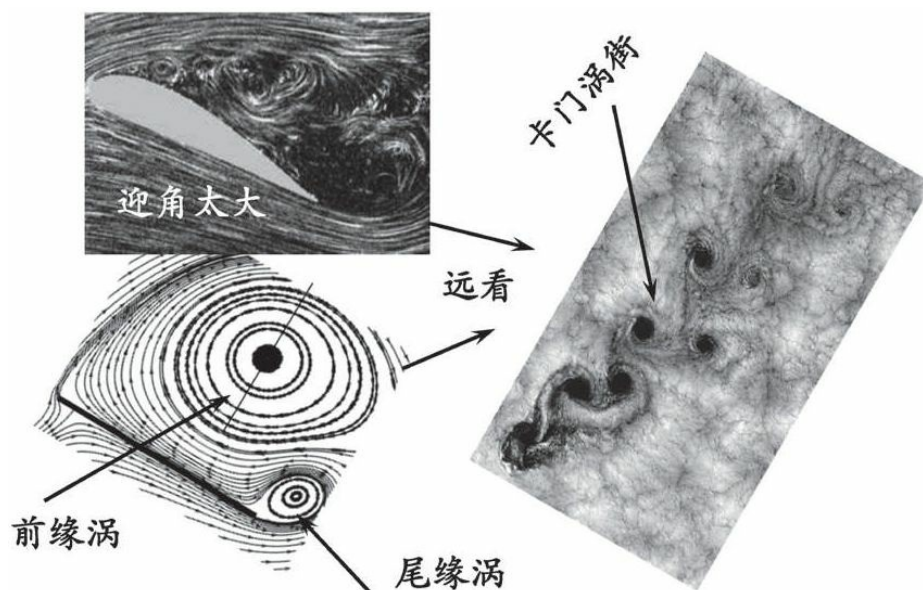


图3.37 卡门涡街：交替产生方向相反的旋涡

电线、桥墩、飞机起落架、建筑物等有气流或水流流过时，都可能产生卡门涡街。一个个旋涡的脱落，引起气压的交替变化。大风天气，我们听到的电线或树枝发出的撕心裂肺的声音，就是这种卡门涡街引起的。

别小看了卡门涡街，卡门还是钱学森的导师，绝对不是小人物。加州理工学院在排名世界第一的那年，在简报中专门把喷气推进、卡门涡街以及地震李氏级别的提出，作为大学三项最伟大的成就。在这三项成就面前，连诺贝尔奖都不值得一提。

剩下的旅途自己探索（图3.38）

机翼的旅行还可以有更多的故事。比如说，鸟翅膀和低速机翼是圆头前缘。可以想象，只有这样，绕过前缘时，拐弯就没那么急，因此就不容易出现拐弯涡。尾缘却又是尖的。只有这样，上下表面过来的气流

才能顺利相会，才能协同地在下面产生正压，上面产生吸力，否则各行其是，就不会有效果。也可以这样理解，由于尾缘是尖的，这样对于有迎角的机翼，就会迫使气流在尾缘相会后统一向下偏转了一个方向。整体上，只有过来的气流被往下吹，才能反过来给机翼产生向上的升力，这就是牛顿作用力与反作用力的基本原理。

迎角太小了，升力不够。迎角太大了，容易出现拐弯涡。拐弯涡把上表面的气流绕乱了，把产生升力的效应给破坏掉了，升力也就小下来了。因此，必然存在一个产生升力的最佳迎角。这个最佳迎角大致在15度左右。

如果飞机超过声速飞行，那么气流在下表面减速后，增加的气压无法以压力波释放到其他地方，因为被超声速气流摠在壁面附近了。对于上表面，由于带迎角，欲急速离开的气流就会试图拉出空隙，这使上表面气流出现膨胀，降低气压。上下表面的气压差也会产生升力。但是，超声速情况下，圆的前缘和上下凸出的形状，容易让空气聚集堵在那里，产生冲击波（因为气压变化以声速传播，在超声速气流中无法向四周有效疏散）。

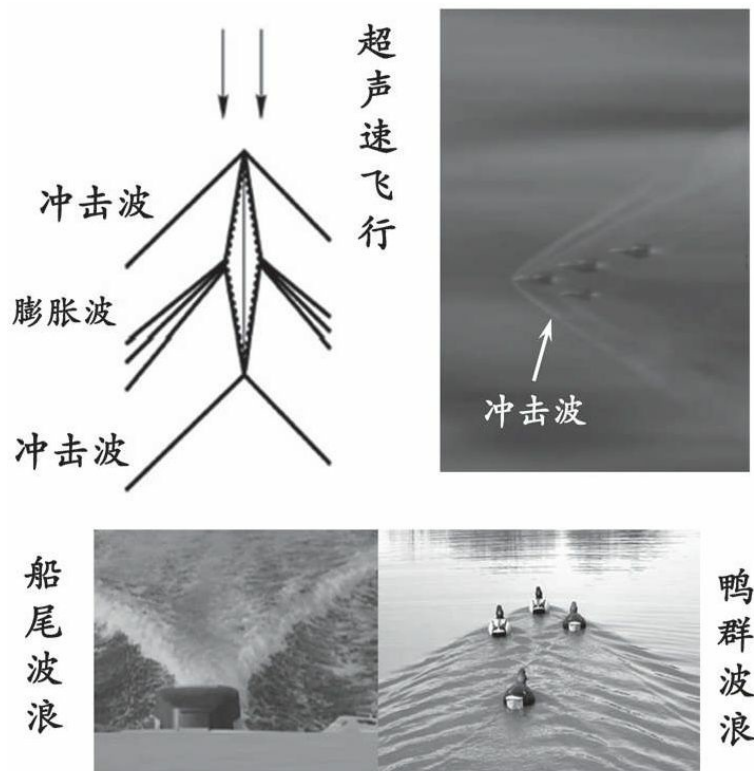


图3.38 空中冲击波与水面开尔文波

低速飞行的机翼，有点弯背，这样产生额外的升力，如同我们能力平时，客气地鞠躬办事会更顺畅。反过来，超声速机翼没有弯度，如同能力超过平均数，可以昂首挺胸发挥作用，再谦虚就反而遇到阻力了。

速度接近声速时，气流速度和声速的大小就旗鼓相当。两强相争，谁也主导不了谁，因此会出现更多的麻烦。因此，必须有一些特殊的措施来应对，由于机身也瞎掺和，因此一些这方面的介绍就留给介绍飞机时再提及。

4. 飞机与发动机

我们终于可以看看飞机了。然而，飞机并不是有了机翼就可以飞

行，还需要稳定与控制，尤其需要发动机提供动力。这会涉及永远写不完的知识。可是，既然我们行在空气中，就不得不在这里挂一漏万地点点飞机和发动机。由于在机翼的气流中旅行就已经够令人疲倦的了，因此这里只简要地介绍固定翼飞机。直升机之类的就不提了。飞机和发动机是一个高技术的家伙，不太适合开它的玩笑，让我们用平淡甚至枯燥的语言来描述一下吧。

固定翼飞机（图3.39、图3.40）

鸟其实最早掌握了飞行奥秘。2014年发现的化石表明，2500万年前的桑氏伪齿鸟，翼展即翅膀展开后的横向长度达6.1~7.4米，而现代飞机歼20翼展是13米。难怪古生物学家丹尼尔·克赛普卡这样说：桑氏伪齿鸟从空中俯冲下来的话会遮天蔽日。

鸟靠扑动翅膀飞行。早期人们模仿鸟，试图也靠扑动机翼来飞行，但一直很难成功。1799年，英国人乔治·克雷在丝绸圆盘上，提出了固定翼飞机的原理：用有独立产生升力的固定机翼，有控制飞机方向的操纵面，有独立产生推力的发动机。因此，机翼、控制面和发动机对于一架固定翼飞机，缺一不可。莱特兄弟1903年发明的现代意义上的飞机，主要得益于他们自制了具有足够强推力的发动机以及发明了控制机构。

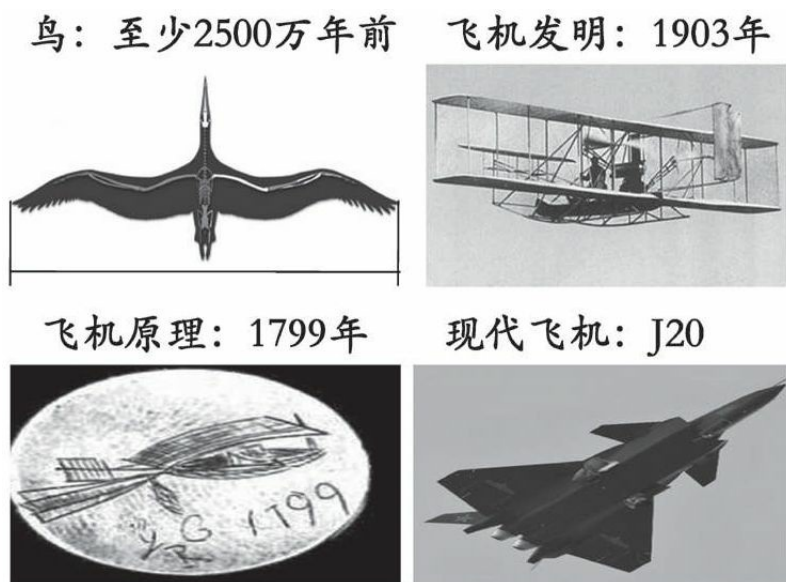


图3.39 飞行的历史

飞机当然有机身，而且往往比较细长，这样才能是流线型，气流容易避让，难以形成较大的压阻。与空气摩擦会产生一定的机身摩擦阻力，也能产生少部分升力。

机翼如同鸟翅膀，呈薄平状向机身两侧延伸。机翼面积足够大，平面形状和翼型（机翼的截断面）尽可能满足产生有效升力的流线型要求，如具有圆的前缘和尖的尾缘、较小的厚度、合适的弯度以及恰当的迎角。机翼除了产生升力，也附带产生摩擦阻力，甚至还有一些压差阻力。

阻力需要靠发动机推力来平衡。发动机推力如果等于阻力，那么就是巡航飞行状态，飞机做匀速水平飞行。如果发动机推力大于阻力，就加速，小于阻力就减速。

升力等于重力时，飞机就能在空中飞行。升力与空气密度、飞行速度的平方以及机翼面积成正比。高空密度越来越低，因此飞得越高，往

往要求越快，或者要求机翼面积越大。

由于机翼的迎角效应，使飞机有很大的升力，比阻力高数倍甚至10倍以上。因此，小的发动机推力，就能让很重的飞机飞起来。



图3.40 飞机飞行原理

尾翼让飞机不会失去平衡（图3.41）

飞机的重力和升力都指向竖直方向，但不在一条线上，因此飞机可能会抬头或低头，失去平衡。就像舞棍，两手横握棍的两个不同位置，朝相反方向用力，棍就转起来了。这种使物体旋转的一对不在一条线上的力产生试图翻滚物体的作用，称为力矩。不能让飞机这样随意翻滚，于是加了水平尾翼。尾翼适当带迎角以产生负升力（或正升力），产生的相反的力矩正好平衡机翼升力和重力不在一条线上的力矩。尾翼很小，但离重心距离远，因此也能产生足够的力矩，这与杠杆原理是一样

的。用一根杠杆之所以能轻易撬动一块大石头，就是因为石头与手离支点距离不一样。

发动机拖着飞机在大气中飞，如果受到扰动后，机尾不听话开小差向左向右偏，垂直尾翼就产生侧向力迫使其摆正。原来，如果出现向左向右的偏转运动，本来顺流的垂直尾翼就与运动方向有了迎角，产生的升力正好使机尾被推回原有的方向。如果机尾向上向下偏转，水平尾翼就会起类似的稳定作用。

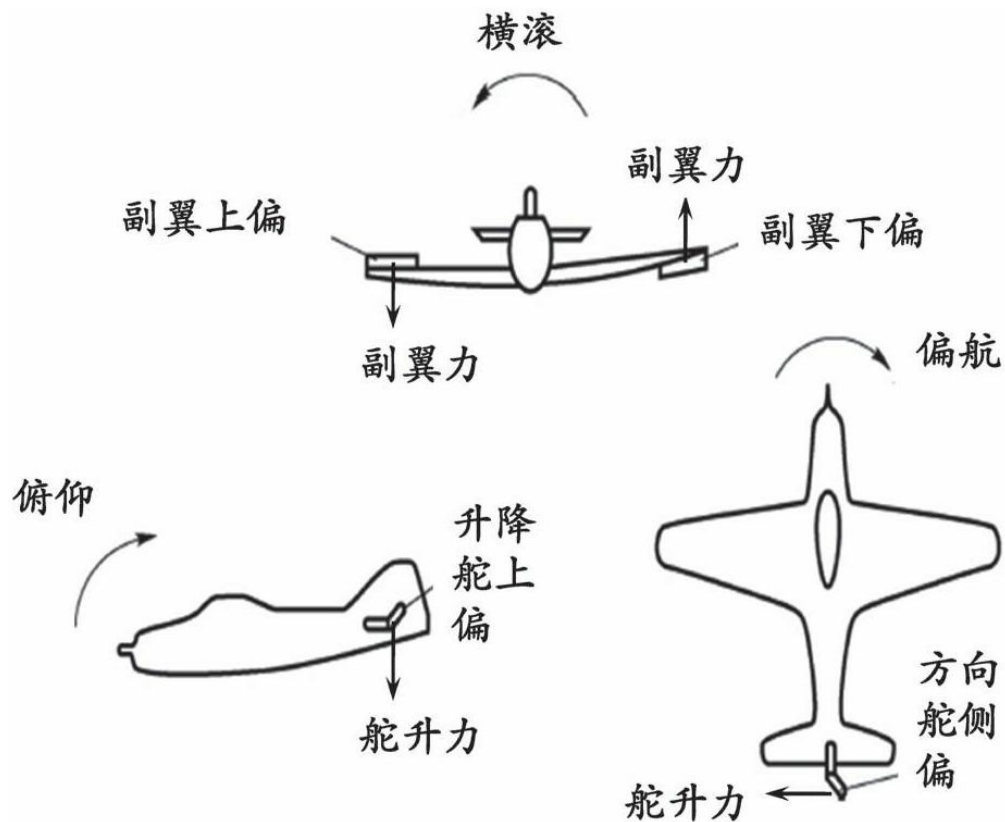


图3.41 飞机副翼、升降舵和方向舵

姿态的改变 起飞着陆

机翼两侧离机身稍远的地方有一对小副翼，可以打开。左右副翼分别向上向下打开，就各自产生了正迎角和负迎角，产生的额外小升力一个朝上一个朝下，转动飞机产生横滚，或使其偏离水平面。副翼产生的力虽小，但距离机身远，于是有了四两拨千斤的杠杆效果。

垂直尾翼上的方向舵的偏转也会产生迎角效应，从而产生侧向升力，使飞机改变航向。尾翼上的升降舵向下向上偏转，产生向下向上的小升力，使飞机低头或抬头。

副翼和方向舵同时起作用，使飞机以偏离水平面的姿态拐弯。偏离水平面，使升力有一个指向拐弯内侧的分量，使拐弯更容易。这如同我们骑自行车拐弯，身体和车倒向哪一侧，往哪一侧拐弯就容易些。

机翼两侧有一对襟翼，靠近机身。起飞着陆阶段，飞机很慢，靠原有的机翼不足以产生能平衡重力的升力（因为升力与速度平方成正比）。抽出襟翼，就能增大机翼面积。另一方面，向下偏转襟翼也能增加弯度和迎角，就能产生足够的升力。

产生动力的秘密（图3.42、图3.43）

发动机据说很难，但基本原理很简单。从本质上说，发动机有受力面。背对着推进方向的受力面（简称背面）如果能增压，就产生推进力。朝着推进方向的受力面（简称正面）如果产生负压，就产生吸力型推进力。侧面摩擦如果朝着推进方向，也产生推进力。有时是三种情况兼而有之。前两种情况也称为压力推进。有时正面也产生正压，但小于背面的正压，还是产生净推进力。人造发动机一般不考虑摩擦推进，但动物在地面靠摩擦推进，动物弹跳则靠压力推进。

以压力推进为例，主要有两种手段产生压差。其中之一就是火箭模式。如果你放过鞭炮，一定见过冲天炮竹，向上冲时后面拖着火焰。这是火药燃烧产生了高压，作用在冲天炮内部，产生了推力。这就是火箭原理。火箭发动机也是这种原理。之二就是叶轮模式，也可以叫升力模式或旋转模式或流动模式。叶片在与推进方向垂直的平面旋转，且旋转叶片相对于旋转平面带有迎角，于是产生迎角升力，这个迎角升力指向推进方向。有的可能是兼具火箭模式和叶片（升力）模式，如涡喷发动机。

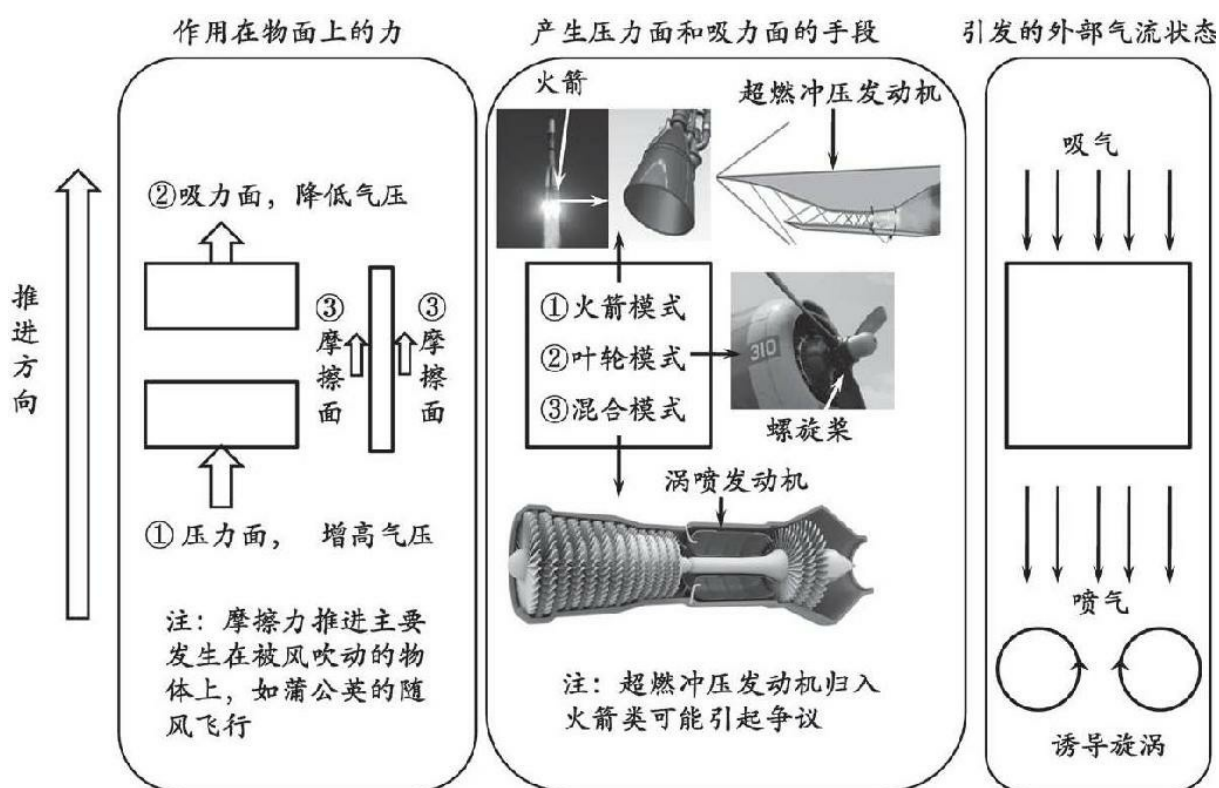


图3.42 从不同角度看推进原理

从发动机诱发的外部气流看，背面出现了喷气，甚至被喷气诱导了旋涡。这种喷气速度甚至旋涡强度可以通过基本的物理学原理与推进力的大小进行等价关联。有时正面出现吸气，吸气速度也可以与推进力的

一部分进行等价关联。但喷气与吸气，甚至被诱发的旋涡，是产生了推进力的结果，却经常被当作了推进的原因。这是一种表述方式上的差别。把结果当原因，只是一种数学物理上的等价。

如果更笼统地表述，也可以把推进原理说成是牛顿作用力与反作用力原理。用针管挤水，用力挤才能把针剂挤出去。被挤出去的液体产生的作用力大小等于用手挤的力，方向相反。用力越大，挤出去的液体飞得越快越远。这就是喷出去的速度越高、推力越大的原理。

以叶片模式（升力模式）中的螺旋桨推进为例。桨叶安装在一个旋转平面内，在不同位置都有一定的安装偏转角，即偏离旋转平面的角度。旋转时，桨叶面与旋转运动的方向有迎角，产生的升力垂直于运动方向，即垂直于旋转平面。于是，这样的迎角升力就作用在推进方向，也就是推力。

离轴心越近，旋转线速度越慢，此时，主要是推进方向决定了气流运动方向，故桨叶在离轴心较近的地方，只是相对于推进方向有一个合适的迎角。靠近桨尖，旋转线速度大，旋转决定了运动方向，故桨叶相对于旋转平面有一个合适的迎角。

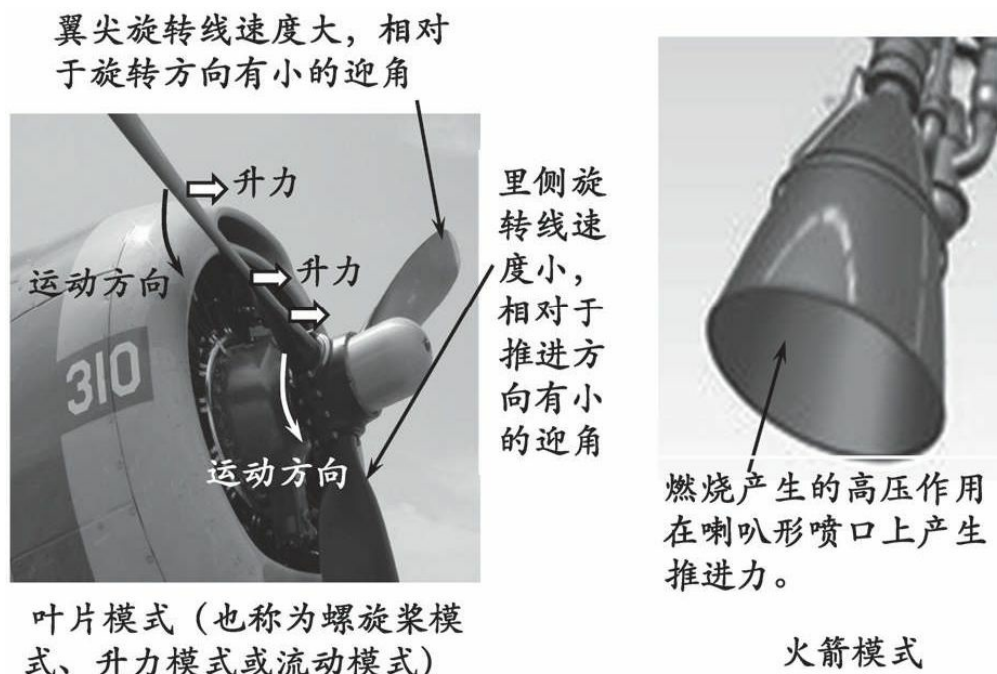


图3.43 产生气压差推进的叶片模式和火箭模式

这种由叶片压力面和吸力面之压差产生的推进，也称为升力模式或流动模式，当然更确切地说就是螺旋桨模式。与前面的火箭模式一起属于两种主要推进模式。

难度极高的喷气发动机则可以形象地理解为既包含了火箭模式，也包含了螺旋桨模式（里面的压气机和涡轮，就是某种形式的螺旋桨，虽然压气机主要用于增压，而涡轮主要用于驱动压气机），虽然一般不这么说。

喷气发动机（图3.44）

喷气发动机不止一种，我们挑最简单的来说明原理。飞机前方的气流通过进气道，先遇上高速旋转的压气机转子叶片。转子有点类似于螺旋桨，只是转子叶片很多。你可以把转子叶片简单看成机翼，只是它们

不是朝飞行方向运动，而是朝旋转平面运动。它们相对于旋转平面有迎角，因此也产生迎角效应，使压力面面向下游方向，吸力面朝上游方向。这种吸力面和压力面的朝向正好使经过一级转子叶片后，气压增高了，当然也适当产生了升力（在推进方向，所以直接贡献推力）。

但一次增高不能满足下游燃烧室燃料燃烧所需要的高压要求。燃烧需要比常温常压条件下高得多的分子热运动。极高速度的分子热运动以及密度足够高的气体才能引起足够频繁和足够强的分子之间的碰撞。化学反应就是通过这种碰撞，将氧气分子与燃料分子通过结合成新的分子，释放能量。

于是需要多级压气机。相邻两级转子之间，安装了不旋转的导流片。被上一级转子带着旋转的气流，遇到不转的导流片，旋转的气流就撞击导流片的压力面（也稍微朝着下游方向），将旋转动能转换成高气压。于是经过导流片气压又增加了一些。

接着往下又是一级转子叶片和导流片，作用和前面一组完全相同。于是，一级一级下去，气压就逐渐增高。有的发动机压气机有十余级，可增压数十倍。具体需要增加多少倍，是不断反复设计和实验总结的规律。这样一边增压，压气机也提供一部分推力。

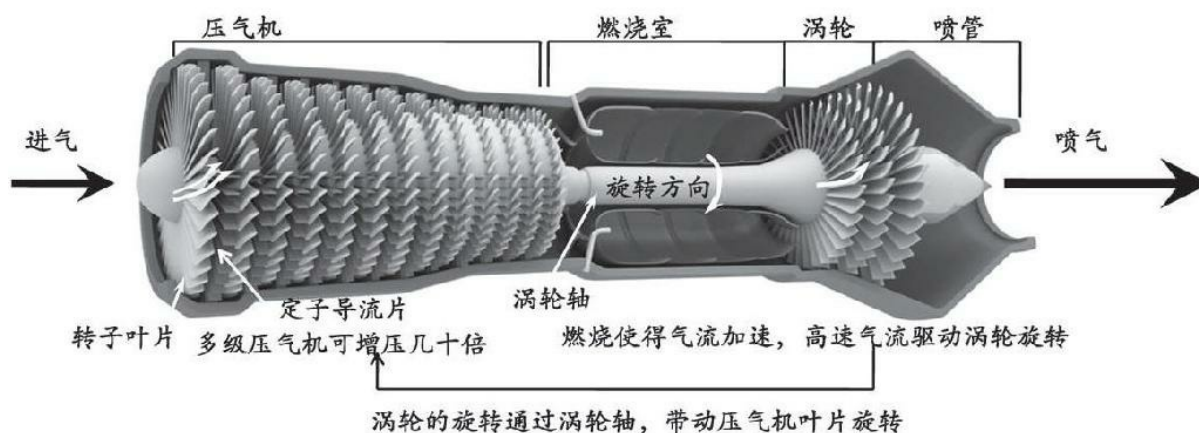


图3.44 涡喷发动机结构与原理示意图

进了燃烧室，气压足够高的气流与燃料混合，就能点火产生化学反应。道理与我们煤气炉烧火差不多，当然要求更高。需要严格选择液体燃料，并且通过喷嘴将其有效雾化，变成极小的颗粒，迅速蒸发成气态，与气流中的氧气尽可能快地混合。点火后，燃料气体分子与氧气分子碰撞，产生化学反应即燃烧，形成高温高压气体，高速向涡轮喷去（也通过燃烧室靠近进气的一侧壁面增压，来提供一部分推力）。

涡轮可以简单想象成风车。风吹风车就转起来了。涡轮叶片也有压力面和吸力面。燃烧室过来的高温高压气流，撞击压力面，就产生了引起旋转的迎角升力（指向周向，驱动旋转）。涡轮的旋转，通过涡轮轴，带动压气机转子叶片旋转。因此压气机转子叶片的旋转是涡轮带动的。

涡轮的压力面朝上游，吸力面朝下游，因此，经过几级涡轮，气压降低了，但一般而言还是比外面的大气压高。从涡轮出来的气流，作用在喷管的尾锥上，又能产生另外一部分推力。

从尾喷管喷出去的气流速度，比进气的速度往往高一倍左右。因此，也可以从火箭模式来理解为何产生了推力。

喷气发动机推力的构成（图3.45）

发动机的推力，按压力的贡献，有时很奇怪。这样看，主要推力不一定全是由发动机提供，前面的进气道和尾喷管也贡献一部分。以超声速喷气发动机为例，如果让超声速气流直接撞击压气机叶片，那样会撞坏，尤其会形成极高的温度。因此得慢慢将气流速度降下来，使到达压

气机的入口时，气流速度已经是亚声速了，比如说正好是声速的一半。

于是需要一个扩张型进气道来将气流速度降下来。进气道之前是前体，相当于收缩管道的一面。在这里，遵循“人遇窄道慢”的道理，气流减速（这种减速还得到冲击波的帮助）并增压。

进入进气道后，已经降为亚声速。进气道是扩张型的，亚声速气流遵循“水遇宽道慢”，于是进一步减速增压。提高了的气压作用在扩张型进气道的内壁上，投影到轴线方向就是推力了。

接着气流进入发动机叶片，故事就和前面一样了。于是可以直接跳转到尾喷管。超声速发动机的尾喷管是扩张型的，这是因为过来的气流比环境气压高，于是作用在扩张型喷管壁上，也形成一部分推力。

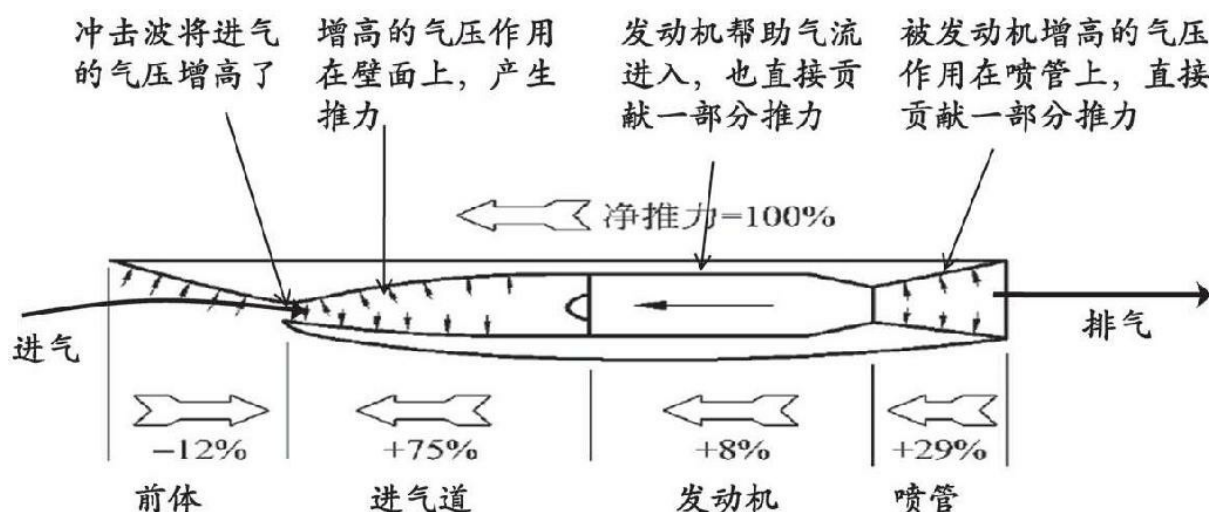


图3.45 喷气发动机A5在马赫2.2时各部件产生的推力比重

著名的A5喷气发动机，在飞行马赫数为2.2时，如果只看气压的作用，进气道产生了75%的推力（进气道前体产生负推力），整个发动机只贡献了8%的推力，喷管贡献了29%的推力。这有点难以理解，难道发动机本身没有必要了，仅仅依靠进气道就能产生75%的推力了？不

是，恰恰是发动机带动了气流有效地进入进气道。如果没有发动机，进气道就进气困难，不会产生推力。

发动机故事续写的困难

喷气发动机几个主要部件的功能是耦合的，合起来就能完成进气、增压、燃烧，燃烧将气流变成高温、高压气流驱动涡轮旋转。旋转的涡轮反过来带动压气机旋转，完成增压。增压了的空气，在燃烧室与喷射的燃料顺利混合与快速燃烧。因此，这是一个闭合循环过程。

压气机的转子叶片不能做得太长，否则增压效果差，叶片长了也容易折断。于是，需要多级叶片，多级压缩。在相邻两级叶片之间，由于气流被上一级叶片转向了，因此到了下一级叶片入口，朝着不利于增压的方向。为了将方向拉直，需要在两级叶片之间安装不旋转的导流片。导流片一方面将气流拉直，另一方面可以将上一级叶片带转了的气流的旋转动能转换成气压。

在高空，大气中的气压一般低于一个大气压，因此注入燃料后不会燃烧。经过压气机或进气道，压力增加到几十个大气压，到了燃烧室，燃料点火后就能燃烧了。燃烧导致气流温度和气压急剧增高，反作用在燃烧室靠近压气机一侧的壁面上，使燃烧室也贡献推力，同时产生高速气流，直接作用在涡轮叶片上，驱动涡轮旋转。因此，为了使涡轮旋转，涡轮就产生了一部分阻力。但这部分阻力与其他部分的推力相比还是很小的。

为了使压气机和涡轮最有效地产生增压和旋转力矩，需要对叶片的形状、大小和安装角进行精密的设计。虽然原理看似简单，但没有一种

简单的理论告诉你要设计成怎样，需要用复杂的理论进行初步设计和分析，找出效率最佳的设计。另一方面，历史经验积累相当重要，需要基于现有的设计，经过不断分析、反复比较和多次实验，才能改进方案。因此发动机的设计需要耗费大量的精力。

除了压气机、燃烧室和喷管的设计需要满足苛刻的要求，对材料的要求也很高。燃烧室与涡轮的温度可能高达1500摄氏度以上，因此用于制造燃烧室和涡轮的材料需要极高的耐高温能力。除此之外，叶片旋转每分钟数千到数万转，使每个叶片交替感受气流的气压变化，这种交替气压变化容易带来疲劳。在一个直径1米左右、长度4米左右的空间内，要安装拥有十级左右叶片的压气机、燃烧室以及涡轮等，因此发动机的制造与安装工艺也相当复杂。如果取每分钟一万两千转，一秒就是200转。半径为0.5米的转子叶片，其翼尖旋转线速度可达每秒624米。这超过了冷空气的声速，会在那里产生冲击波。冲击波一方面导致很大的阻抗，另一方面会带来其他损失。因此，转慢了不行（增压不够），转快了也不行（产生有害的冲击波）。

航空发动机要求耗油率要低，要求安全可靠，要求可维护，要求低噪声。民用航空发动机平均寿命得一万小时量级。

于是，航空发动机被称为飞机的心脏，甚至属于反映工业实力的工业之花。

接近音速的飞机（图3.46）

鸟的翅膀和低速飞机的机翼带有弯度，向上凸起。机翼也相当于机身的凸起物。这种凸起拦住了从侧面避让的气流流道，因此得加速才能

通过。

这里，飞行马赫数本身就接近1，上述加速就很容易导致气流超过声速。站在飞机上看，前方本来是亚声速来流，现在局部出现超声速区了，因此称为跨声速流动。

一旦在凸起的部位出现超声速气流，就得按“人遇宽道变快”的原理，进一步加速了。于是，凸起部位的下游气压越来越低，在凸起物下游产生吸力，形成额外的阻力，也称为跨声速阻力。

本来气流速度和声波速度就势均力敌，互不相让，你还在那里添堵，不把矛盾激化才怪。

这种阻力非常大，导致不加适当改进的飞机很难接近和突破声速，这就是音障的来源。局部超声速气流撞击下游的空气，后者无法及时避让，因此形成伞一样的冲击波。由于这种冲击波的存在，人们也把这种跨声速阻力归结于冲击波，把阻力增加归结于跨声速波阻。

既然是局部凸起的堵塞引起了跨声速波阻，那么把机翼的上表面削平（称为超临界机翼设计），让机身在链接机翼的位置瘦腰（称为跨声速面积律设计）以及让机翼向后掠（如同燕子的翅膀一样向后掠），就会有效地减弱这种凸起效应，减小波阻和冲击波的强度。这样，飞机就轻易突破了音障。

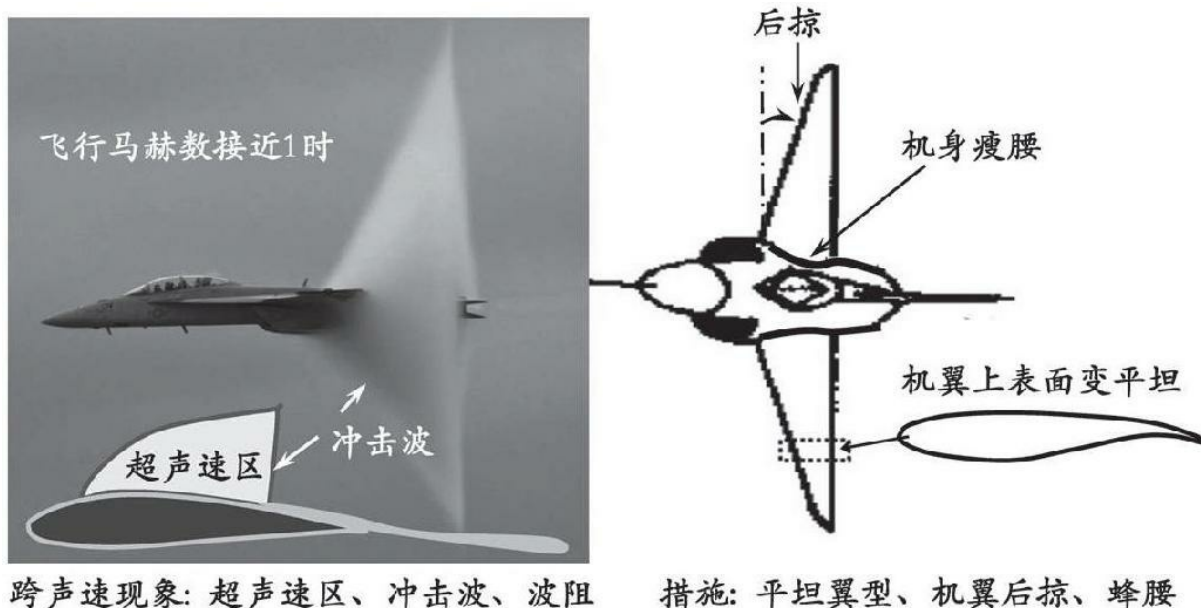


图3.46 跨声速现象及其减弱的措施

当然，凡事有个度，以蜂腰为例，蜂腰后虽然显得更美且减弱了跨声速波阻，但瘦得太厉害也不行哦。有的飞机腰瘦不下去，就把前段机身加粗点，这样子看上去就显得蜂腰了。这样子也能减弱跨声速波阻。这么简单的道理本来谁都懂。

5. 翼尖翻转涡 飞机起降频率 候鸟的迁徙

飞机在云端飞行，或者在湿度较大的低空飞行，我们会看到下游有两道旋转方向相反的旋涡。那就是翼尖翻转涡。这种翼尖翻转涡对下游飞机有不好的影响，但聪明的候鸟却巧妙地利用翼尖涡，借力飞行。翼尖涡让两个翼尖涡之间的气流下降，而外侧的气流上升。排队的候鸟就驾驭在翼尖涡外侧上升的气流之上，它们真得感谢飞在最前面的鸟。

飞机的翼尖涡（图3.47）

产生升力的机翼的压力面（下表面）气压高，吸力面（上表面）气压低，因此，在翼尖的位置，下表面的高压就会驱动气流从翼尖侧边绕到上表面，这种气流翻转实际上就是一种旋涡，旋涡吹到下游，于是拖出两道长长的旋涡，就是翼尖涡。

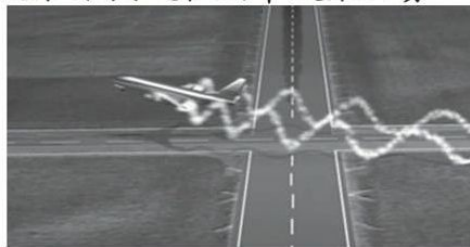
从飞机尾部向头部方向看，左侧的翼尖涡顺时针旋转，右侧的逆时针旋转。如果下游远端有云彩，那么就可以清晰地看到翼尖涡将云朵卷成螺旋状。由于翼尖涡旋转产生离心力，会迫使涡心气压降低。如果当地气象条件使空气中的湿气接近饱和状态，那么在涡心低气压作用下水蒸气就会凝结，让我们看到白雾般的翼尖涡。

翼尖涡那种翻转方向使在两个翼尖涡之间产生一股向下吹的气流。这会使下游的其他飞机处于危险状态。因此，机场航班起降时，前后两架飞机必须隔开足够长的时间或距离，以便前一架飞机的翼尖涡在摩擦等因素作用下耗散得差不多了。

被翼尖涡卷曲的云彩



饱和水蒸气在涡中凝聚白雾



飞机襟翼打开时的旋涡



候鸟利用翼尖涡排队迁徙



图3.47 翼尖涡与候鸟迁徙

飞机越大，翼尖涡越严重，起降间隔就需要越长。因此，飞机不一定做得越大越好。

国际民用航空组织（ICAO）对飞机起降制定了标准，将飞机分为重型、中型和轻型三类，按先后起降的飞机的类型规定了飞机起降间隔时间或距离，时间间隔大致在80秒至180秒之间。前面一架飞机起飞后，下一航班需要等几分钟才能起飞，就是这个道理。

飞机起降阶段，速度还没有提升到巡航时的速度，因此，与速度平方成正比的升力不够。为了获得能平衡重力的足够升力，起降阶段需要打开襟翼。襟翼从机翼尾缘伸出来，并且向下偏转，等效于增大了机翼面积和机翼弯度，产生足够的升力。然而，在襟翼的翼尖，也会有类似于主机翼翼尖涡的旋涡产生。

除了对下游飞机起降有不利影响，翼尖涡对自身也有影响。翼尖涡迫使两翼翼尖涡之间产生向下的气流，作用在机翼上，相当于改变了前方过来的气流的方向，使气流向下偏转了。这样，与这种偏转了的气流垂直的升力，有一个分量就指向飞行相反的方向，即产生了一种额外的阻力。为了减弱这种额外阻力，人们发明了翼梢小翼，即在翼尖竖立一个小机翼，从而把翼尖涡减弱以及抬高。抬高后，翼尖涡离机翼就远了一些，负面作用当然就减弱了。

翼尖涡在两翼翼尖涡之间（内侧）产生下降气流，当然在翼尖外侧就产生上升气流。如果处在下降气流中不利，那么处在上升气流中就有利。“二战”中盟军将轰炸机以V字形编队飞行，使下游飞机处在上游飞机翼尖涡外侧的上升气流之中，结果发现最高可节省18%的燃油。

候鸟迁徙与人字形飞行（图3.47）

秃鹭之类的候鸟远距离迁徙时，通过排成V字形或人字形，下游鸟利用上游鸟翼尖涡外侧的上升气流来获得额外升力以及推力，减少体力消耗。然而，生物学家不满足于这种显而易见的道理。2014年，华盛顿大学生物系的莫伊杰里斯（Florian T. Muijres）和狄更生（Michael H. Dickinson）在《自然》周刊上报道了从飞机上观测秃鹭作V字形飞行的细节。他们发现，候鸟在不断调整身体位置和翅膀扑动频率，以便最有效地节省体力。

候鸟迁徙有时可能利用了大气环流。灰鹱，一种乌黑的海鸥，从夏季末离开加利福尼亚飞回新西兰，这一旅行飞行的距离累计达到惊人的3万9千英里，是有记录的最长的距离。

英国《生物通信》2015年报道称，一种体重只有12克左右（大致12张名片的重量）的林莺可在两到三天的62小时内，不间断、不吃不喝地飞行2540千米左右的距离。这对应的对地速度平均约为每秒11米。如果林莺不借助风，而是在空气中飞行，那么为了克服阻力（比如说3克的阻力），飞行62小时需要消耗15600多卡的热量。假如林莺没吃东西，完全靠消耗自己的肉变成克服阻力的能量，那么需要消耗782个左右自己的身体（12克鸡肉，有20卡左右的热量），才能自主（而不是借助风）飞这么远的距离。这显然是不可能的。

但据观察，林莺在迁徙前，先吃饱昆虫，使自己增重一倍。12克猪肉含有23500多卡热量。假设1克昆虫含的热量与1克猪肉差不多，那么林莺吃了12克昆虫，可以支撑飞越那么远的距离而不借助风力。

事实上，大气中是有风的，在10千米左右的高度，平均风速可以达

到20米每秒。如果林莺在迁徙时，能飞到这样的高度，靠这样的风，也能在三天内飞5000千米左右。

那么，到底是吃得饱吃得好，使林莺能扑动翅膀飞那么远，还是需要悬在大风中，让风吹那么远？

3.4 自然界的运动：一曲动感的旋律

聪明的鸟和昆虫通过交替改变翅膀方向，使迎角升力一部分用于平衡重力，一部分形成推力。一些植物的种子插上了翅膀，像直升机一样飞翔。被风吹落的蒲公英种子会举起冠毛伞。下落的雨滴会通过变形来降低下落速度。动物的腿既像倒挂的单摆也像弹簧，奔跑时也就是摆动和振动。不要以为肌肉没什么事，肌肉同时起到发动机、刹车、弹簧和支撑等作用。鱼儿摆摆尾巴就往前游，有的也弯曲身体再发力，如同我们赛跑前的预备姿势。有的鱼用胸鳍划桨或像螺旋桨一样产生推力。当然，还有雪花。草木之花多五出，独雪花六出，雨滴落水咚一声，雪花落水有铃音。

1. 鸟与昆虫的飞行秘密

很容易想象，昆虫因为翅膀太小而产生不了足够的升力，就像太小的飞机也飞不起来一样。可是它们就是能飞起来，而且能快速机动，能驾驭风吹雨打。鸟比昆虫大些，比飞机小些，它们能折中昆虫和人造飞机的一些飞行技巧。不仅如此，鸟与昆虫的翅膀既能产生升力，还能产生推力和控制力。原来，它们知道用某种频率和方向扇动翅膀，既把翅膀当机翼又把翅膀当螺旋桨，既能扇出旋涡又能把旋涡当气球使。

鸟飞行的秘密（图3.48、图3.49）

天空飞鸟拥有一对羽扇般的翅膀。这对翅膀是完美的流线型外形，既有合适的迎角也有恰到好处的弯度，从而在平飞时当然能像飞机机翼

一样产生升力。

升力好解决，不够就把翅膀长大点、把迎角调合适点、让机翼形状完美点。

可是，鸟不能总是靠大气上升气流升上蓝天，然后在那里翱翔吧。它也得向前飞行，于是必然会有摩擦阻力。它没有发动机，推力从哪里来？再说，逃生和捕食，追逐伙伴，也需要加速嘛，这更需要推力。

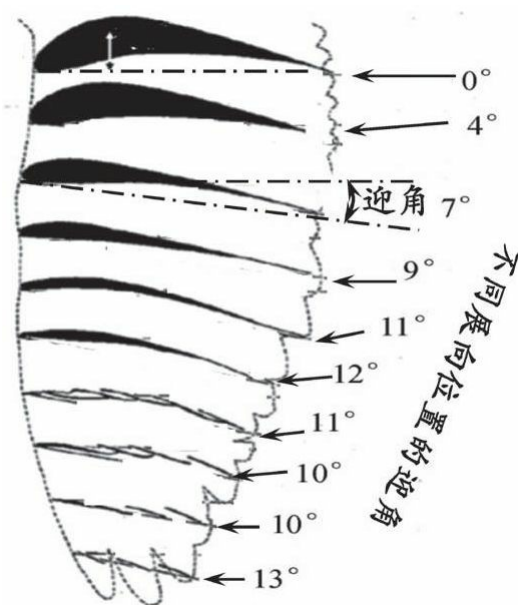


图3.48 既有迎角也有弯度的鸟翅膀

鸟说，连这都不能解决，俺还是鸟吗？不就是缺产生推力的螺旋桨吗？俺抬高翅膀并调整出螺旋桨那样的姿态，向下扑动一下，就相当于螺旋桨转了小半圈了。当然，扑下去后，还要继续抬起，以便接着下扑，一次下扑产生的推进力只够飞行一点点时间的。

向上抬起时不相当于螺旋桨反转了？反转产生的不是推进力，而是回拉力。一推一拉，平均而言就什么都没了。可是，鸟也是有脑子的。向下扑动时伸展开翅膀，推进力足够大。向上抬起时收缩翅膀，这样就不会有什么回拉力。

靠近身体的内侧翅膀，旋转线速度不怎么大，因此螺旋桨效应就不怎么强。因此，内侧翅膀往往不怎么扑动，让那部分做固定翼飞行，如同飞机机翼一样。外侧翅膀做扑翼运动，产生推进力。

让一侧翅膀拍动幅度比另外一侧大一点，那么左右翅膀产生的力的

大小和方向就不一样了，这显然可用于拐弯甚至作出一些复杂的机动动作。

发懒时，翅膀完全展开不做任何扑动，利用惯性做滑翔或利用上升气流做翱翔。

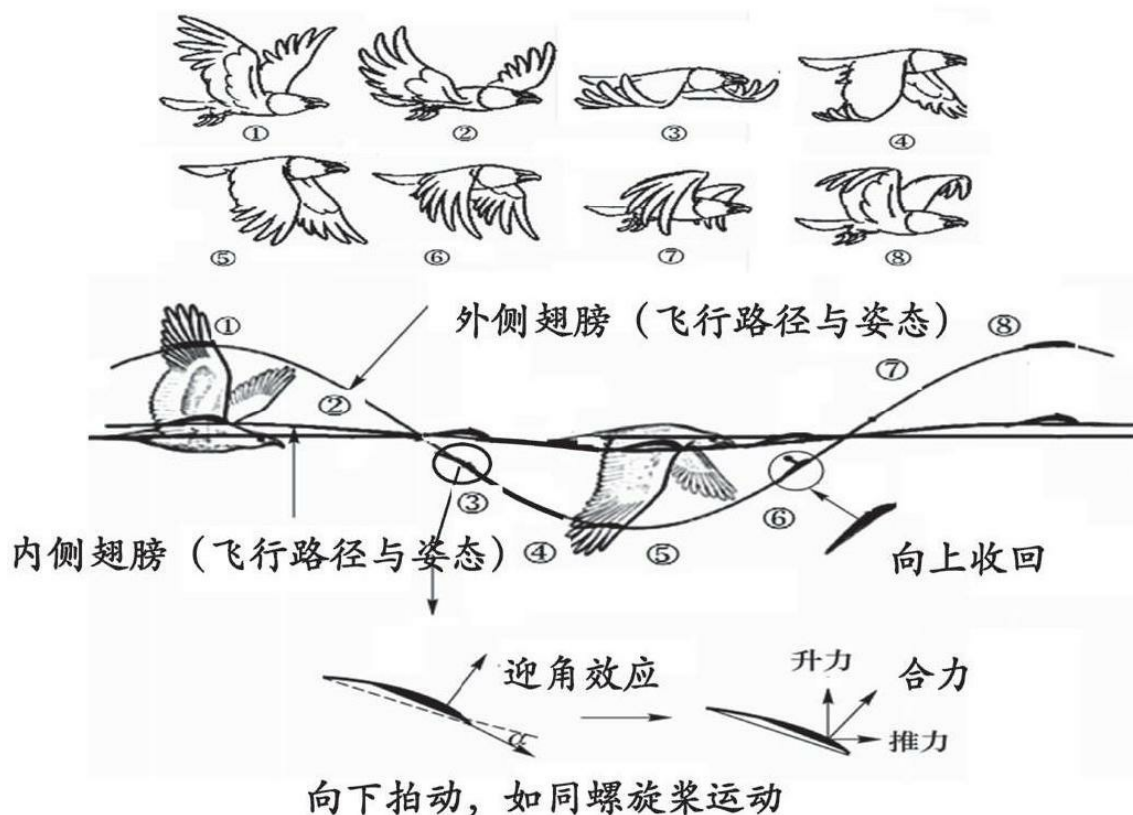


图3.49 鸟的翅膀运动与升力和推进力的产生

昆虫飞行的秘密（图3.50）

昆虫那么小，能产生足够的升力吗？你可以很简易地理解升力应该不够的道理。升力与面积以及翅膀转速平方（从而长度平方）成正比，重量与体积成正比。两者的比值当然与长度成正比。昆虫这么小，长度

就小，于是升力就应该比重量小多了，按理飞不起来了。

既然能飞起来，就有道理。但道理也得依靠现代摄像机技术才能搞清楚。昆虫扇动翅膀的频率每秒可达数百次。例如，蚊子翅膀扑动频率高达每秒500次以上，一种称为蠓的小黑蚊每秒可拍动翅膀1046次。这么快扑动翅膀，凭肉眼哪能看清。不搞清楚翅膀运动，就没法研究能产生足够升力的原因。科学家大智若愚，居然不会让昆虫饿几天或者将其拍晕，让翅膀慢慢地扇动，这样不就看得清了？

可是，即使动物保护协会还顾不得保护昆虫，科学家也会注意形象，不会采用这种不人道的方式对待昆虫。再说了，以这种方式取得结果，论文肯定发表不出来。

有了现代摄像技术，就搞清楚翅膀形状和拍动方式了。原来，已经不像鸟的翅膀，昆虫翅膀是薄片型翅膀，前缘也是尖的，没有弯度。翅膀在一个平面内来回扇动。特别像船橹来回摆动的情況，也不知古人是否真的采用饿晕了的昆虫，通过观察翅膀慢拍后受到了启发，才发明了橹。

向前扇动时，翅膀抬得高高的，迎角很大。扇到最前面，就翻转翅膀，抬起角度向后扇动。通过翅膀翻转，使向前向后扇动时，都带正的迎角，产生升力。这个迎角最大时，可以达到35度左右。

可是，不是说了，翅膀前缘太尖和迎角太大会产生拐弯涡，按理不利于产生升力。鸟翅膀的迎角就不会超过15度，35度时迎角升力效应就很小了。

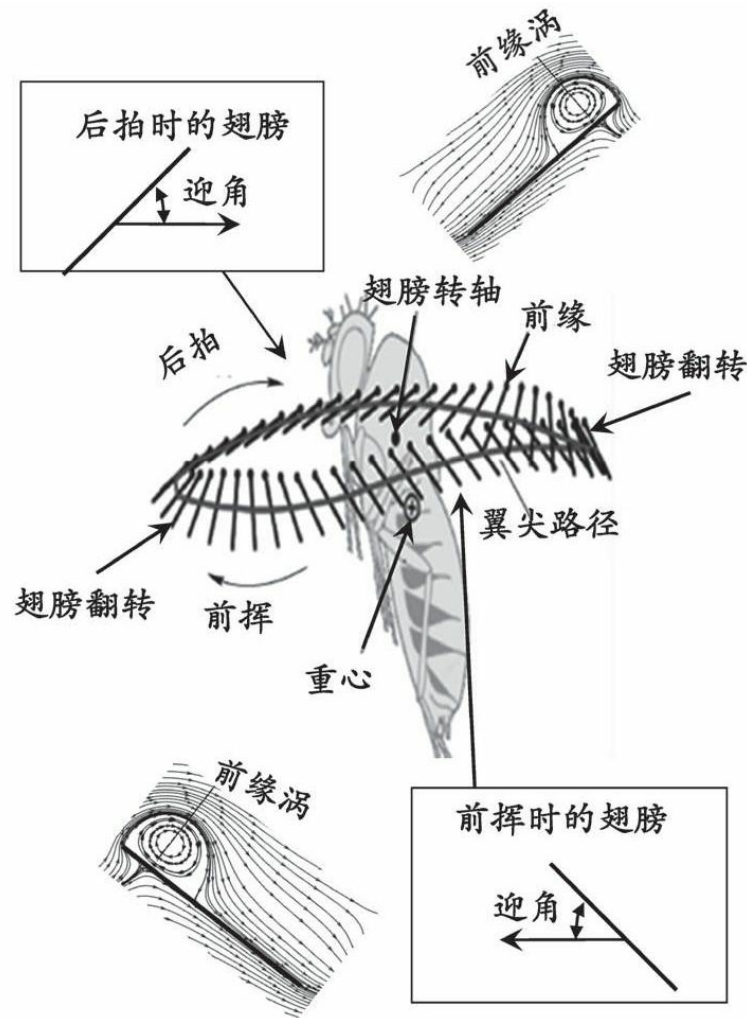


图3.50 昆虫翅膀反转、前挥和后拍运动

原来，拐弯涡也不是一无是处。它离开前缘之前，就像吊在翅膀上的一个气球，会提供很大的力拉着昆虫。这是因为，拐弯涡的涡心气压低，作用在翅膀上，就提供了很大吸力。只有前缘涡脱落后，升力才会下掉，就像悬吊重物的气球吊绳断了一样。

昆虫很聪明，前缘涡脱落时，翅膀已经翻转，改变了运动方向，进入新一轮产生前缘涡和高升力的过程。你这个前缘涡没用了，我就改变方向，产生新的前缘涡去了。

因此，昆虫巧妙地利用了翅膀上方前缘涡脱落之前的吸力作用，而当前缘涡脱落、升力快要急速下降时，翅膀已经反转了，与前面的前缘涡没什么关系了，又开始利用新的前缘涡了。

将翅膀扇动平面放在水平面上，昆虫就能悬停。将身体以及扇动翅膀的平面向前倾斜，升力的一部分就变成向前飞的推力了。翅膀的一侧抬高点儿，另一侧放低点儿，升力就有了一个转弯的分量，因此就可以轻易转弯了。

昆虫令人不可思议地掌握了这些科学家费了很大劲儿才搞清楚的原理，能驾轻就熟地悬停、前飞、倒飞、俯冲。反应快，转弯半径小。悬停时，在空中不动，如直升机一样。

2. 蒲公英与降落伞 蝴蝶果与直升机

飘落的种子、树叶、纸片、雪花等，没有什么去主动控制它们。它们无法控制自己，只能与空气共舞。可是，蒲公英会巧妙地举起冠毛伞，枫树果会插上螺旋桨，舍不得掉落在地面。也不知道是它们模仿了人造飞行物，还是人造飞行物模仿了它们。总之，都想在空中留得久点儿，飞得远点儿。它们既可能是下落的物体，也可能是飞行的物体，因此是飞行与下落的一种混合状态。

举着冠毛的蒲公英种子（图3.51、图3.52）

奥杰布瓦族有个传说：南风与蒲公英相爱了。这说明，蒲公英可以随风飘扬。在约瑟夫·安东尼的《蒲公英种子，伟大的梦想》中，是这样描述蒲公英种子的：她拥有完美的生命，美丽的飞行姿态，能在风暴

中幸存，可忍受黑暗的煎熬，从不放弃，是自然中最伟大的成功故事之一。于是，蒲公英成了勇气、耐心与不屈不挠的象征。

蒲公英种子由瘦果和若干毛茸茸的冠毛伞组成。数目不等（如20多个）的冠毛伞如同插在莲座上。冠毛有数十根直径为几十微米的蛛丝状细软毛（比人的头发丝还细）。

在无风的情况下落下时，瘦果朝上，冠毛伞朝下，一边下落一边翻一个跟头，在5厘米左右的距离内翻完，使冠毛朝上。接着像降落伞一样徐徐下落。下落速度在3厘米左右的距离内，从0达到每秒0.5米左右的的最大速度，接着掉落到每秒0.2米左右的速度。这个最终的速度，就是向上举起了降落伞后徐徐下落的速度。

五类播种方式：风吹，水流，动物，炸开，人工播种

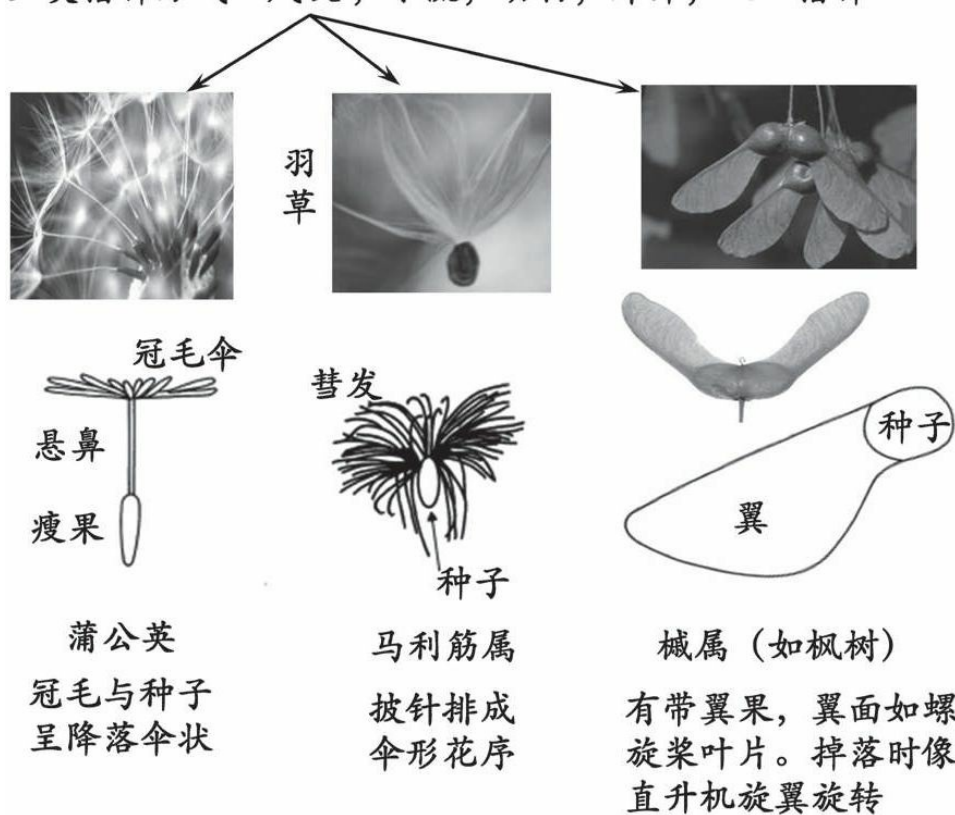


图3.51 随风播撒的种子

这样的下落速度，从0.2米左右的高度掉下来，没有风的话，1秒左右就落地了。因此，蒲公英种子需要有风，才能飘得更远，配得上100米以外的长距离播种的名声。可是，得有每秒100米速度的风，才能在落地之前飘100米远。如果只有水平风，这也吹不了多远。于是，人们认为，是大气中垂直向上的对流风，让蒲公英一下子落不下来，随风飘得很远。

如果你看到一粒蒲公英在头顶漂浮，表明那里的垂直对流正好维持它的高度。你抓住它，把它的种子摘掉，扔向空中。它没了种子，变轻了，会一下子蹿到高空。

冠毛中的软细毛，很容易被风吹着走。原来，这些软细毛被风吹时，会受到空气的摩擦力。这个摩擦力作用在风的方向，带着蒲公英走。摩擦力与物体的表面积成正比。蒲公英这么多软毛既增大了总的表面积，又能柔软变形，使软毛顺着风的方向将摩擦力最大化。那么为什么不将软毛变得更细，使摩擦力更大呢？因为太细后，就容易断了。

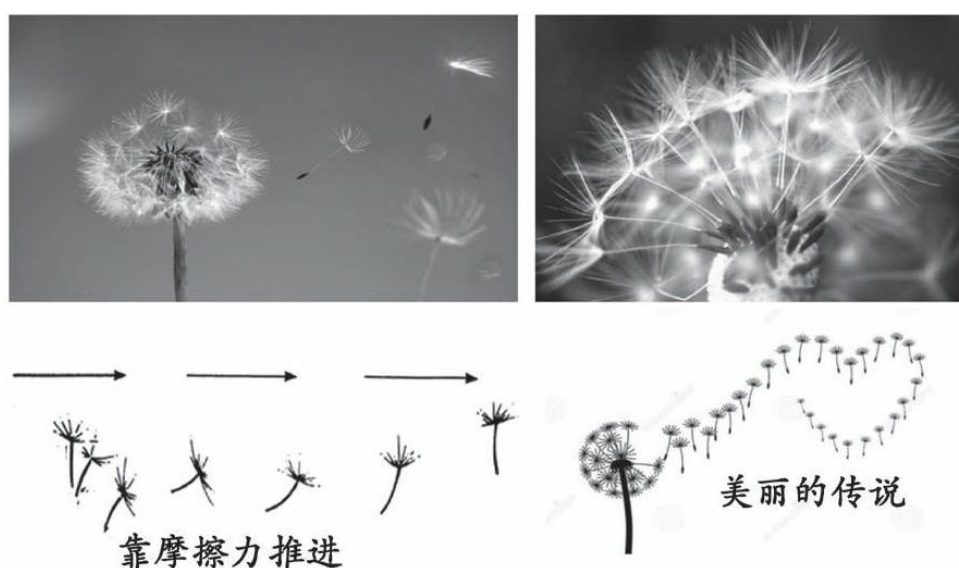


图3.52 蒲公英种子举着冠毛伞随风传播

可以做一个简单计算。一根圆柱如果等分成100根更细的圆柱，那么侧面的总表面积会增加10倍。如果将蒲公英的瘦果卸下，冠毛伞在风的作用下，会飞得很高很高。

种子在重力作用下下落，下落速度如果太快，那么冠毛伞就起到一些降落伞的作用。只要遇到一股向上的风，冠毛伞就在摩擦力作用下，举起蒲公英种子飘得更高。低的坠落速度与对风的跟随性好，使蒲公英种子可以在水平和垂直气流作用下，传播很远。

终端速度（图3.53）

中学时代就会说，落体的速度与大小无关。十大关于思想的科学实验中，有伽利略的重力实验。亚里士多德曾认为，自由落体速度取决于物体的质量，他认为质量越大下落越快。伽利略于是构造了一个简单的思想实验来反驳。按照亚里士多德的逻辑，那么一个轻的物体和一个重的物体用绳子绑在一起从塔上丢下，由于重的物体下落的速度更快，两物体之间的绳子会被拉直。此时，轻的物体对重物会产生一个拉力，使重物体的下落速度变慢。如果是这样，从另一角度来看，两个物体被绑在一起以后的质量应该比单个物体的质量都大，从而被绑在一起的物体比单个物体下落更快。从两个角度按亚里士多德的理论来思考问题却得出十分矛盾的结论。伽利略以这个思想实验证明了亚里士多德的理论是错误的。因此，伽利略认为，自由落体的下落速度与质量无关。

但这是指的自由落体。物体从空中下落，空气会提供风阻。一般情况下，还真的是物体越重下落越快。物体越小越轻，下落速度越慢。对于小的颗粒和液滴，还会有一个终端速度，即重力与风阻取得平衡的速度。小颗粒和液滴最终以这个终端速度下落。

一粒小球的阻力与迎风面积以及速度平方成正比，从而与半径的平方成正比。重力与体积成正比，因此与半径的立方成正比。因此，重力与阻力的比值与半径及速度平方成正比。如果小球半径非常小，那么稍有速度，阻力就远大于重力，从而阻止有过快的速度。这使越小的物体终端速度越小。

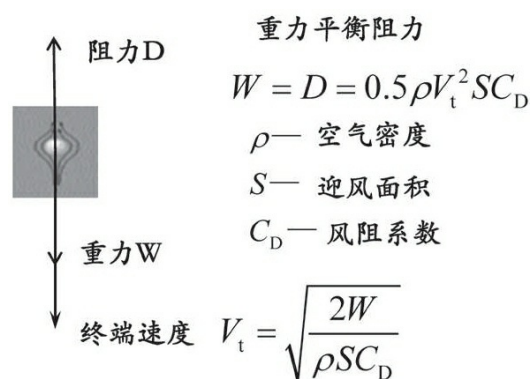


图3.53 下落物体的终端速度

大气中一些小颗粒直径在微米量级。它们下落得更慢。水蒸气超过一定的浓度，在低温作用下，凝结而成直径在微米量级的小水滴（多半在2微米至15微米之间）。天气预报所指的雾，是在这种水滴浓度多到将能见度将低到1千米以下时的气象现象。污染物中也有一些颗粒的尺寸是微米量级。

直径为数微米的雾滴不会像直径为数毫米的雨滴那样在重力作用下快速掉下来。如果考虑重力作用下的下落，那么在摩擦阻力作用下，直径为10微米左右的雾滴的终端下落速度每秒只有不到3毫米，一小时下落10米左右。在空气分子热运动的撞击下可能使下降速度更慢，使局部的雾滴不会一下子靠引力下落而迅速消失。

尘霾（包括雾霾）中的细微颗粒的直径也在微米量级上下，终端下落速度极小，因此会悬浮在大气中。如果地面温度比高空的高，那么这种温差引起的上升气流就会将局部高浓度雾霾颗粒带入高空，消散在大气中。如果反过来，地面温度更低，即形成逆温（高空温度更高），那么就很难产生上升的气流，小颗粒污染物既无法被带到高空，又由于终端速度太低无法及时落到地面。

如果不是小球，而是柱状物体，情况类似。细小的物体终端速度会很小。蒲公英的种子带有很多细毛，每一根毛都起到使终端速度极小的作用。因此种子终端速度极小。

降落伞与翼装飞毯（图3.54）

降落伞的形状如同雨伞，着陆阶段伞面铺展开，凸的一面朝上，凹的一面对着下方。降落伞垂直下降时，伞圈所在的平面完全迎着下降方向，迎面撞击空气，形成高压，反作用在伞盖上，减缓下落速度。

一般是小物体下落时才有终端下落速度。但降落伞这样的大物体，由于阻力大，也有终端速度。为了尽快下落，希望下落速度越快越好。为了避免受伤，希望越慢越好。伞盖的面积越大，阻力也越大，最后选定的速度是避免受伤的最大速度，可以依据这个速度设计伞盖的面积。

单纯的伞降运动，约85%的受伤发生在着陆时。人从距地面1.5米高处跳下，落地方式对的话，一般不会受伤。按自由落体运动计算，对应的落地速度大致是5.5米每秒左右。由此不难理解跳伞选手着陆时的速度限制。一般情况下，着陆速度不要超过5米每秒，防护措施、体质和经验好的选手可以承受更大的着陆速度（指垂直速度）。除此之外，选手落地时还可能有一定的水平速度。将下降速度设为安全落地速度，就可以反算出降落伞展开后的面积。



图3.54 蝙蝠飞行与蝙蝠翼装飞行

蝙蝠背着一对涡环，产生高升力，比同样大小的普通翅膀的升力高数倍，可以达到普通机翼的4.8左右。不仅如此，蝙蝠的翅膀具有多达24个骨骼，并且覆盖了弹性好的皮肤。这使蝙蝠的翅膀具有柔软性和弹性，能自适应地调节气流，让它最有效地产生和控制升力，可以在空中悬停和在飞行中快速转弯，掉头所需距离仅是其身体长度的一半。

有一种极限运动，称为翼装飞行对接。带降落伞包的选手从距离地面数千米的直升机（直升机飞行高度很难超过6千米）一跃而下，经过精准的时差后打开降落伞，站在悬崖边上的翼装选手通过助跑起跳，展开蝙蝠一样的翼装。两人在空中分别飞行一段时间后，通过调整姿态，拉近距离，以“人体飞毯”的形式完成对接，持续“飞毯状态”一段时间后分离，各自展开降落伞着陆。

翼装飞毯借用了蝙蝠翅膀产生高升力的形状。飞行运动员可以通过双臂和双腿的调整，控制身体在空中缓慢滑翔，同时调整航向。翼装飞行滑翔每下降一米的同时前进约三米。无动力翼装飞行进入理想飞行状态后，飞行时速通常可达到每小时160~200千米。

摇着螺旋桨的翼果（图3.55）



图3.55 带翼果的下落和空中飘行

如果在起风的枫树林中见到小“直升机”在空气中旋转，不要奇怪，那可能是松树果子在旋转。

蝴蝶树、枫树、某些松树、紫葳属树等种子或果实带有“翅膀”，使种子能像飘落的名片那样产生升力，可以翻滚地斜着飘落；或者产生螺旋运动，使下落获得的旋转速度抵消了直线下落速度，从而下落更慢。无论何种情况，种子的掉落更多来自于大风的作用。由于下落更慢，甚至由于产生升力而飞起来，因此在风的作用下飘得更远，甚至飞越山丘、河谷，将种子播撒到更远的地方，生根发芽。

枫树果长在树上时，带着两片螺旋桨一样的翅膀，具有明显的弧度（即弯度）。每片翼有一颗种子，两颗种子在一起，使带翼果造型颇像两片旋翼的直升机。掉落时两个翼会旋转起来，就像直升机的旋翼。落地后两颗种子可能会分开。有的带翼果只有一片翼。单片翼的种子下落时，也会以螺旋形式旋转。枫树果有旋转玩具、直升机和旋转球等别名。

欧洲的枫叶种子只有一片翼，从种子的一侧长出一叶翅膀。翅膀的长度与种子的重量是如此匹配，以致这种种子在下落时，也会自旋起来。由于这种自旋，下落就很慢，即使在微风条件下，也能飘往很远的地方。

3. 雨滴下落 流星自焚 航天飞机与返回舱再入

大气中可能有雨滴、冰雹等下落。幸亏空气阻力将它们减速。不仅如此，雨滴太大的话，空气还会将其变形或者撞碎，让它们无法获得砸伤人的下落速度。流星再入时，初始速度就极快，大气就将其烧毁。大气的力量比想象的大，连航天飞机和返回舱那样的庞然大物以极高的轨道速度再入时，也能把它们的速度降下来。当然，既不能降得太慢也不能降得太快，因此航天飞机和返回舱有自己的下降走廊。弹道导弹既可

以采用滑翔方式再入，还可以采用“打水漂”方式一弹一弹地再入。

雨滴下落与碰撞（图3.56、图3.57）

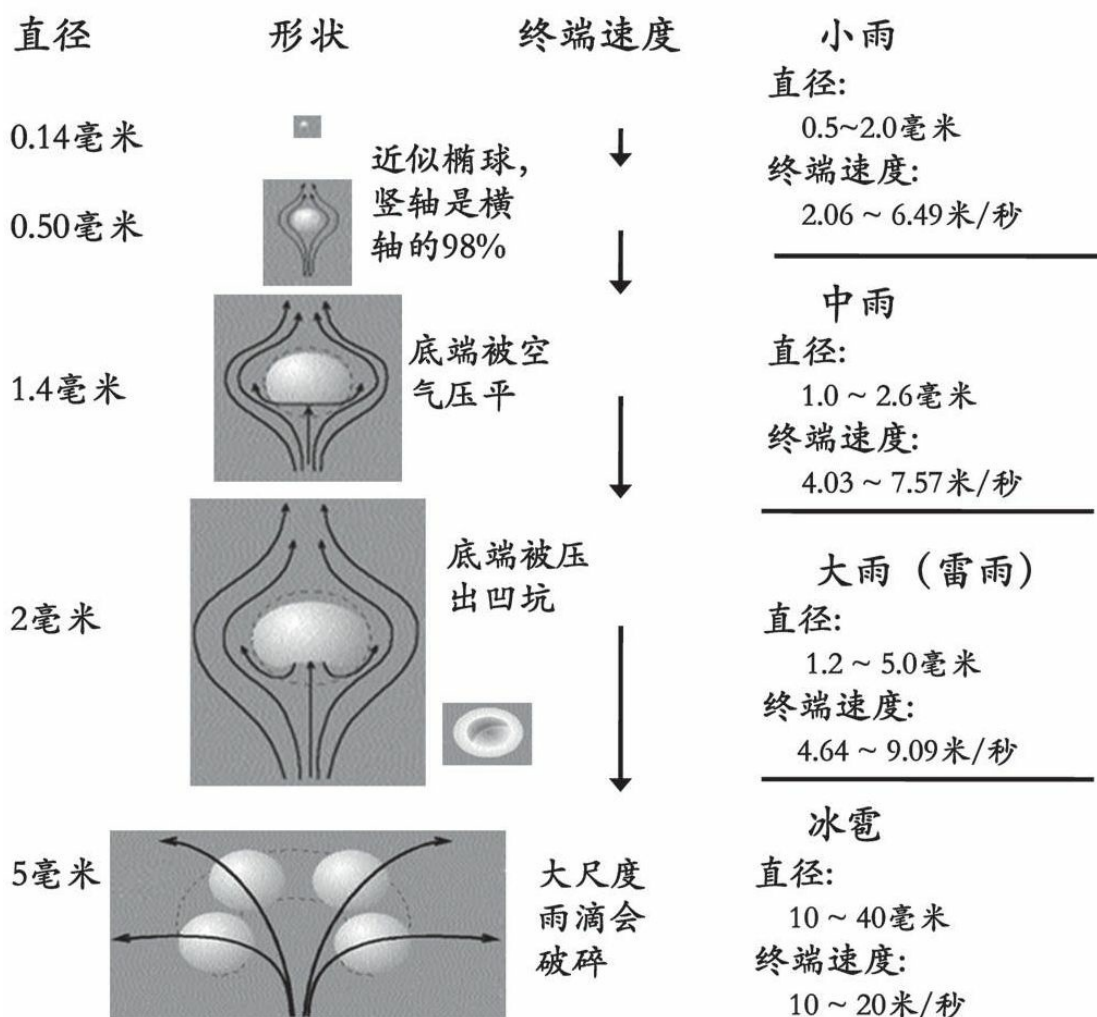


图3.56 雨滴下落与终端速度

积雨云最高高度可达8~12千米，即接近对流层顶了。如果空气不给雨滴阻力，那么雨滴从12千米高度下落到地面，按自由落体理论，以每秒获得9.8米/秒的新速度不断加速，可以算出大致需要50秒到达地面，落地速度达到每秒485米左右，比音速还快。这么快的速度，差不

多和子弹一样，会把脑袋砸出窟窿。如果这样，地球就不会进化出我们这样脆弱不经撞击的人了。

原来，雨滴在空气阻力作用下，也会很快获得不太大的终端下落速度。雨滴的大小并不是固定的一个值，小的直径有0.2毫米，大的直径超过6毫米。有的可能达到1厘米。冰雹的直径可达4厘米。事实上，风阻系数也与速度和雨滴尺寸有关，且与雨滴的形状有关。如果雨滴尺寸达到5毫米或以上，那么在与空气发生作用时，会出现大幅度变形而碎裂成小雨滴，就像海洋在大风中会出现碎浪一样。直径为2毫米的雨滴下落时，底部会变形出现酒窝型凹坑。这是因为，正对下落方向的点气压最高，容易被挤变形或者挤碎。

气象学家史蒂文·豪斯特迈耶（Steven L. Horstmeyer）总结了雨滴下落的速度以及其形状的变化。实测的雨滴终端速度大概在每秒几米左右。依据雨滴大小，终端下落速度一般在每秒10米以下。

如果是一粒直径为几毫米的小石子，以每秒数米的速度垂直砸在头面上，还是很难受的。然而，雨滴是水，砸中物体后，会流动起来，会变形。依据尺寸大小、撞击速度和表面湿润程度的不同，可能会出现铺展黏附、飞溅、碎裂和回弹几种情况。

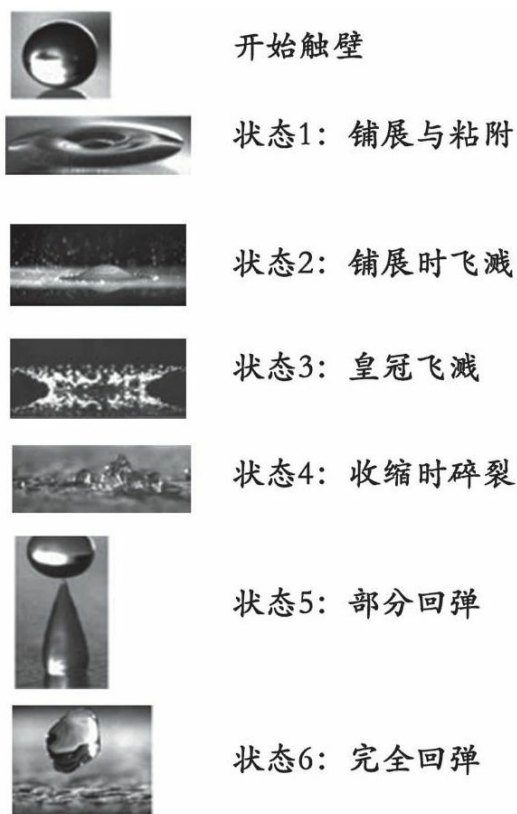


图3.57 液滴与壁面碰撞时的形态

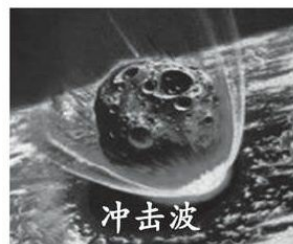
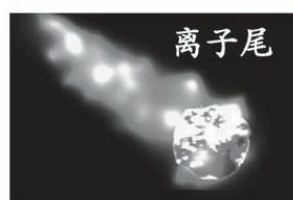
雨滴等液滴撞击固体，在惯性和重力作用下，会展开成碟状。与此同时，表面张力的回拉作用使碟状液体收缩。如果撞击较快且液滴较大，那么固壁的反作用压力会将碟片状液体举起来，形成皇冠一样的形状。表面张力作用在液体边缘还容易包裹出小液滴，形成碎裂现象。如果液滴较小从而表面张力相对而言较大，那么表面张力会一直包住液滴，让其像乒乓球一样可以回弹。

雨滴打在干燥的人体上，一般会黏附、碎裂或飞溅。就是说，雨滴的动能不会完全变成打击力，而是一部分通过变形铺展碎裂消耗掉了。

流星体与陨石（图3.58）

尺寸在数厘米或以上的太空流星体以每秒十几千米甚至几十千米的速度再入大气层，就会出现流星（夜间我们能看见这样的流星）现象。有的会在高空发生爆炸，尺寸合适的会拖着一团火球再入，甚至发出巨大的声音。那些尺寸在一拳头或以上的，可能没被烧干净，落到地面。落到地面的称为陨石。

虽然称为陨石，实际上除了密度在3~3.5吨每立方米的石质陨石外，还有密度在7.5~8.0吨每立方米的铁质的陨石，以及石质和铁质混合的陨石。



流星高速再入，冲击波加热气流，气流进一步用摩擦加热物体。热到连电子都被打出来了，形成离子组成的火焰

能在烈火中生存下来的，到了地面就称为陨石了

图3.58 流星高速再入时被烧起来了

考虑一块半径20厘米左右的球状陨石，按空气阻力与重力平衡，可估算出终端速度为每秒200米以上。实际上，在更高的高度，空气稀薄，密度比地面空气小多了，没有足够的距离或时间降低初始再入速度，形成每秒200米这样的终端速度，从而落地速度可能比这大。这么快的速度落地，当然十分危险。好在能落到地面的尺寸足够大的陨石，十分稀少。

原来，流星在再入过程中，以每秒数十千米以上的速度与空气发生摩擦，会产生极高的温度，大部分流星在高空就烧得差不多了。

如果是圆球形状的流星体以比音速高许多倍的速度再入，到了低层大气，头部会出现冲击波，预先加热空气，可以达到数千度。在这么高温下，被加热的空气甚至出现电离，这种由离子和自由电子共同存在的高温气体，也称为等离子体。

每年再入大气层的太空流星体和其他碎片有一万多吨。虽然大气层对那么高速的这种物体的减速效应不强，但通过冲击波与摩擦，将大部分烧干净了，落到地面的数量很少，尺寸很小。

高速再入飞行器的外形 大自然的灵感（图 3.59）

大气摩擦能烧坏流星，航天飞机和返回舱那样的庞然大物以轨道速度再入时，不能让它们烧坏。科学家与工程师当年为此绞尽脑汁。总是会出现能工巧匠，能用启发性思维从大自然吸收灵感，以致能诞生像超声速客机以及航天飞机那样的高速飞行物。

与飞机在大气层平飞不同，卫星、航天飞机和返回舱等航天器是在

高于100千米的某条轨道上绕地球运动，有点类似于月球绕地球运动。它们再入大气层时，拥有的初始速度依然是轨道速度，至少每秒7.9千米。这么高的速度下来，要想安全地落到地面，必须靠巨大的空气阻力减速。另一方面，速度降下来之前，摩擦加热会试图烧坏它们。因此，对于航天飞机、返回舱以及弹道导弹，再入过程中既要不被烧坏，也要降低速度。通过从大自然和其他直觉启发灵感，人们设计出了合适的外形以及再入模式，使任务得以完成。而流星和卫星碎片等，则希望在高空即被烧干净。你可以想象流星带着一团火球再入，多么壮观。

人们在研究高超声速弹道导弹时，首先想到将导弹头部做成尖的，以减弱压力波堆积即冲击波带来的波阻。可是实验发现，这样很快就被气流加热烧坏了。头部越尖，越容易烧坏。

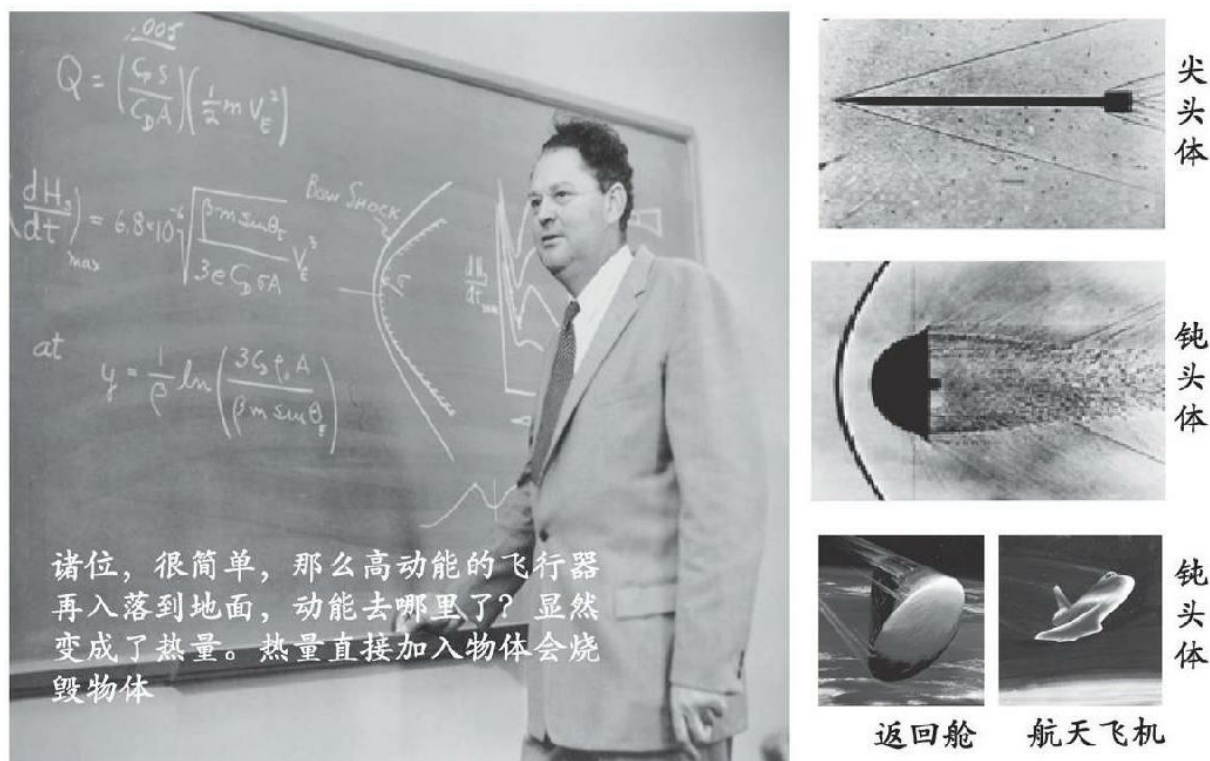


图3.59 钝头推出冲击波减弱气动加热

能落到地面的陨石都是非流线型钝型物体，有的甚至像球形。似乎只有这种形状，才不被烧干净。于是，自然不难想到，高超声速导弹以及以高超声速再入的航天飞机和返回舱，都应该从陨石吸取灵感，要做成钝头的，才能避免被烧毁。

阿龙（Allen）为此提出用钝头体代替尖头，解决了当时遇到的热障问题，被评为美国当年最杰出的科学家。可见，从大自然找灵感，往往能出其不意地解决重大科学与技术问题。他的道理很简单：钝头体以超过音速的速度飞行或再入时，头部会产生冲击波。冲击波一方面将气流加热，另一方面将气流推动到与飞行器速度差不多。被加热的气流从边上流走，不会直接去加热物体。被冲击波减少了相对速度的气流，摩擦力从而摩擦加热也减小了。这个理论看似简单，可解决了当时洲际导弹的设计问题，因此是划时代的贡献。道理知道后就很简单，但第一个能想到的才伟大。

航天飞机和返回舱之类的做成钝头既减弱了加热，又能获得很大的阻力。它们再入时，就是需要利用很大的阻力来减速。因此，一举两得。

人类正在研究在30千米左右高空做水平飞行的高超声速飞机，为了防热需要做成钝头的，为了减小波阻需要做成扁平的。这种矛盾提高了这类设想的难度，其实现依赖技术的进一步突破。

再入走廊与弹道（图3.60）

设想返回舱从100千米高度直接垂直落下来。这会导致几十倍的加速度（减速度）。人完全无法承受这么大的加速度，飞船结构也无法承

受。

因此，垂直再入存在加速度太大的危险（称为过载太大）。只能近似水平地绕着地球斜着再入。但斜得太厉害，减速太慢，又会甩离地球。

于是，航天器的再入路径有限制，能成功安全再入的称为再入走廊。太过于沿垂直方向容易出现危险，太水平则会甩出去。

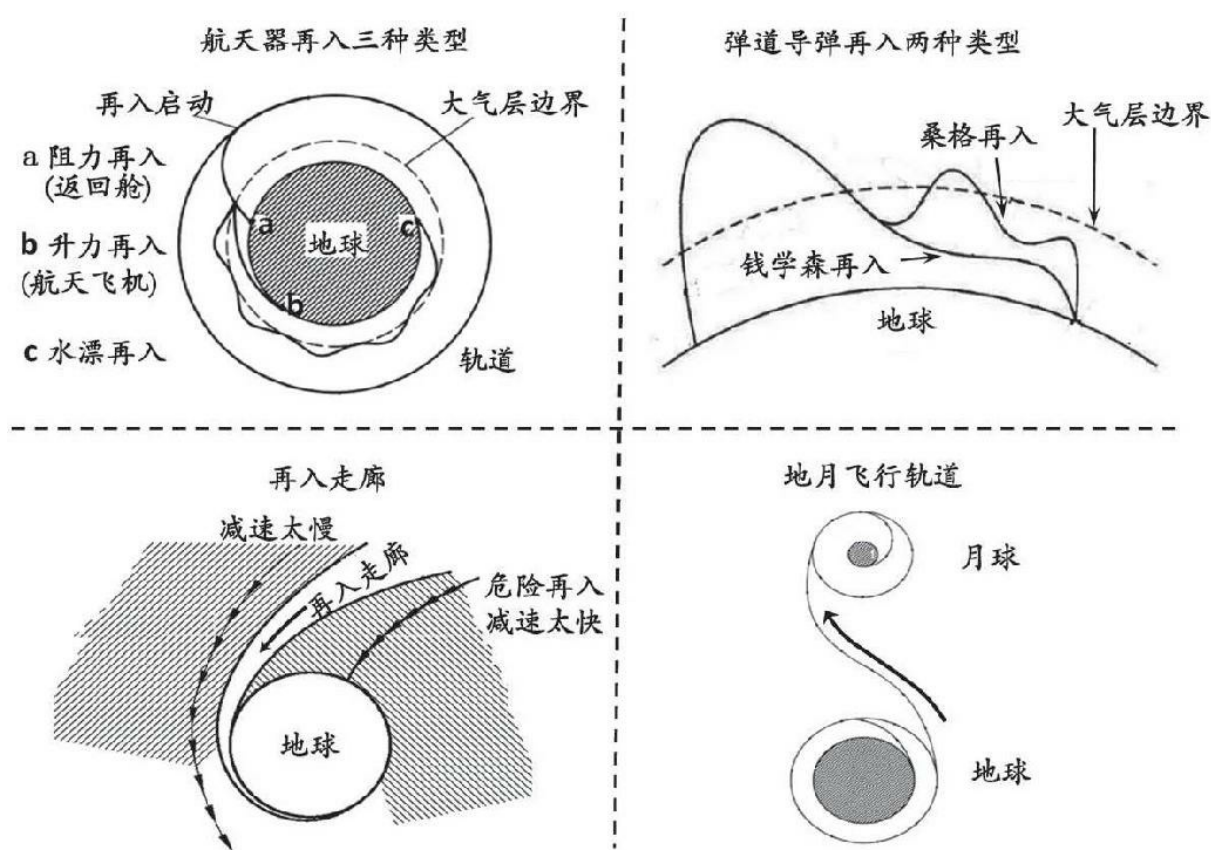


图3.60 再入走廊与地月飞行轨道

航天飞机带有机翼，因此可以部分利用水平飞行的升力，来减缓再入速度。还有一种再入方式，就是让飞行器在大气层边缘打水漂一样波浪形运动一段距离，择机再入。弹道导弹再入也是类似的情况。钱学森

设计的再入模式看上去是在滑翔，而桑格设计的再入是水漂型，即波浪形再入。

4. 生物推进的相似性 鱼的游动

空中的飞鸟、昆虫和蒲公英等，采用的推进方式与水中的鱼类以及地面动物推进方式相比，没有本质区别。因此，这里欣赏一些更多生物的推进方式，虽然有的不在空中。动物没有发动机，只能靠自身的附属肢体（如翅、鳍、腿、尾等）和身体的分段运动来产生推力。动物推进无论多么复杂，其核心原理非常简单：靠身体上某些组织给环境的空气、水或陆地施加一个力，环境就给动物施加反作用力，这个反作用力推动动物前进。动物本身并不懂得推进原理，它们在迁徙、逐食、躲避、求偶或其他需求的驱动下，经过自适应的进化，本能地掌握了推进技术。甚至连肌肉都可以起到发动机作用，除此之外肌肉还有刹车、弹簧和支撑等作用。

人力推进（图3.61）

我们当然更熟悉自己的推进行为。这些行为主要发生在陆地行走和奔跑中，当然对于少部分人，也发生在水面和水下。要在空中，恐怕只能借助飞行工具。

在陆地行走或奔跑，要靠两条腿（用手爬山和攀岩就不说了）。这里面蛮有学问的。行走时，一条腿如同一个倒悬的单摆。跨步过程中，后腿会向前倾，于是人的身体会下沉一点点。前腿落地后逐渐变得竖直，身体就抬高了。每条腿在跨步过程中，有点儿像撑杆，还有点儿像弹簧。奔跑的过程中，身体高度的起伏更明显。高低起伏将重力转换为

一部分能量。掌握得好的，可以使身体高度的变化转换成的动能再转换为行走或奔跑的动能。如果是短跑比赛，甩手的方式恰当也能起到类似作用。左脚向前迈时，向后甩左手，右侧类似。这样可以加快迈步的速度。

因此，走路和跑步的姿势可以分解为倒单摆运动和弹簧运动，原来的运动往往是简谐运动的叠加。

人在游泳时，主要靠手掌手臂和脚掌推水或划水产生反作用力。踩水的一种方法，是往下踩水时尽量将小腿向两侧展开，足底放平，增加产生水阻力的迎水面积，回抽时尽量收拢，减少回抽时的迎水面积（减少阻力）。这样，下踩时，阻力朝上，回抽时，朝下的阻力较小，平均下来就能产生向上的力。另一种方法是把足部当机翼，利用足底是平的、足背是凸的这种弯度效应，通过让平直的脚步在一个平行于水面的平面内画圆，利用足底足背的弯度效应产生向上的升力。经过适当训练，就可能利用这些踩水方式或其他自己熟悉的方式，产生向上的力，与浮力一起将人托起来。

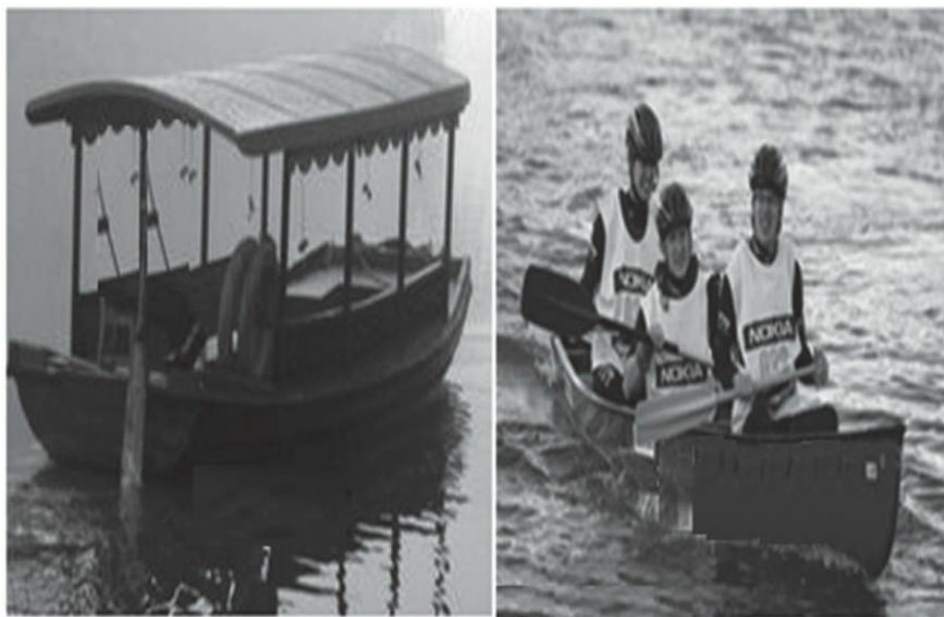


图3.61 驱动船航行的橹和桨

用人力驱动船的行驶也涉及推进。常见的划船工具是双手操作的竹篙撑船、龙舟用的双手划水的划桨、固定于立柱上的两侧同时划水的棹，以及在船尾用于摇动的橹。当然，纤夫用纤绳拉船也是一种推进方式。

生物推进（图3.62）

生物推进表面上十分复杂，但水中生物采用的基本原理可以归纳为阻力模式、升力模式以及火箭模式。阻力模式可能令人误解，因为我们意识中的推进是为了克服阻力的。所谓的阻力模式，是躯体或者附属肢体向后运动（即朝生物运动方向相反的方向运动）时，产生的阻力变成生物本身的推进力。踩水往下蹬，在脚掌上产生了与蹲腿方向相反的阻力，但这个阻力沿着让人浮起的方向，属于想要获得的推进力。

按推进部位的运动形式分，又可以归类为划桨模式、扭动（波浪）模式、旋转（扑翼）模式和喷水模式。无论何种模式，只需要存在迎着前进方向的吸力面或背着前进方向的压力面，或者两种面都存在，就能产生推进力。

首先说一下鱼的推进模式。通常人们会认为鱼是通过摆动鱼尾即尾鳍来获得推力。但事实上，鱼产生推力的方式很多。尾鳍摆动往往只提供了部分推力。其产生推力的原因与划桨时顶水运动类似。当尾鳍摆在左侧或右侧时，突然回摆到中间，相当于在顶水运动，产生的阻力指向鱼的游动方向。

鱼的身体往往也在左右摆动，从上方俯视一条向前游的鱼，其身体

在向左右方两侧摆动时，有时身体弯成C字，有时像波浪。波纹形态向鱼尾方向传播（但躯体并不向鱼尾方向运动），因此就形成压力面（身体向后推水的那一面）和吸力面（身体背水的那一面），产生推进力。小蝌蚪就是通过舞动小尾巴产生推力。小尾巴抖出波形，形态向后传播，交替产生迎水的压力面与背水的吸力面，以此产生推进力。一些极小的生物带有鞭毛，像蛇一样舞动鞭毛产生推进力。

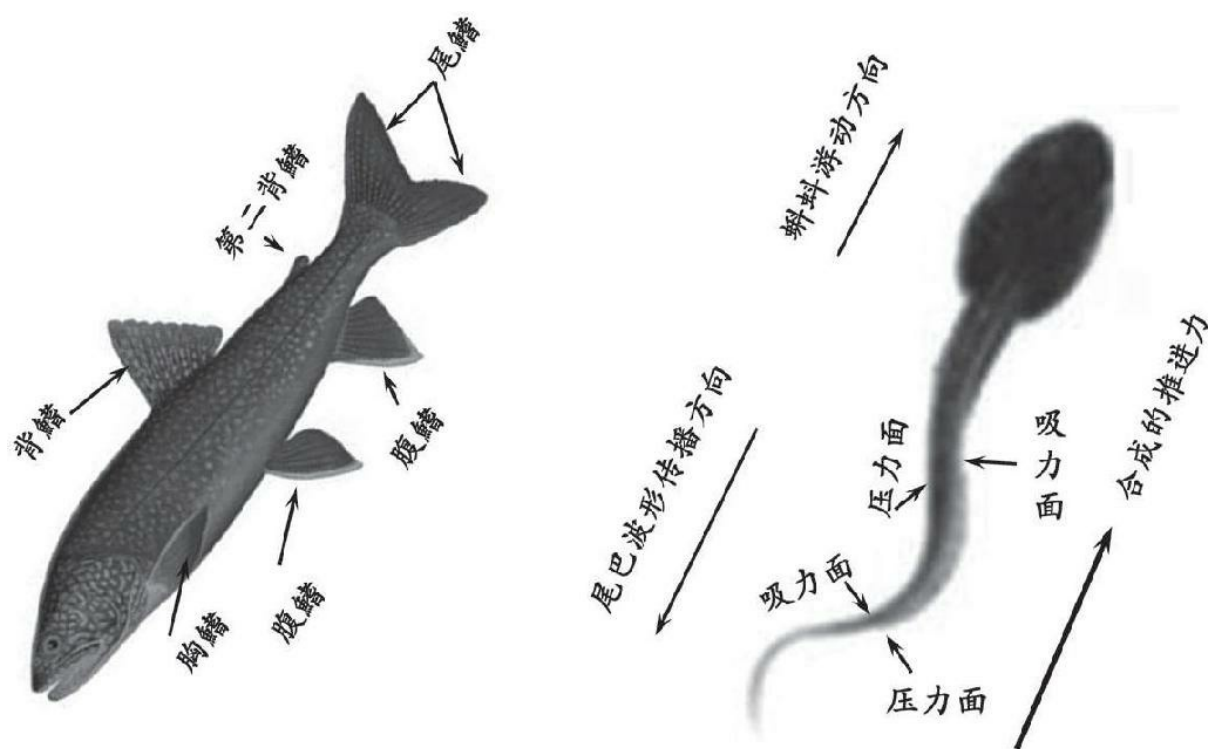


图3.62 鱼和蝌蚪的推进肢体

也有一些鱼靠胸前的胸鳍，用划桨模式或扑翼模式驱动。划桨时，胸鳍的平面首先尽量朝鱼尾方向的垂直方向，向后划水时迎水面积大，产生足够的阻力（即推进力）。复位时则将胸鳍平面尽量与运动方向平行，这样阻力就小，不会抵消划水时形成的有用推进力。如果胸鳍平面对着前方，像昆虫扑翼那样运动，那么就产生升力。这个升力指向游动方向，也就是推进力。

有的水母呈子弹形状，有的呈碟状，通过喷射水柱产生推进力，这是一种火箭模式。

总之，细看每一种生物的运动与推进方式都很特别，但遵循的原理就那么多，甚至可以这么归纳：通过附属肢体（如翅、鳍、腿、尾等）或身体的一些部分做相对运动，以推动环境的水、气体或陆地，产生阻力、升力等，这些力指向推进方向，就是推进力。做这些相对运动时，通过改变姿态（迎角）、迎水迎风面积等，使有效推力得以保证，抵消推力的力尽量减弱。

推进中的普适原理（图3.63）

加州大学伯克利分校集成生物学系的迪金森等六名科学家，在2000年的《科学》杂志上，发表了一篇有关动物推进的综述性文章，几乎涉及所有陆地、水里和空中的动物推进过程附属肢体和身体各部分的运动方式，以及产生推进的原理。

通过全面分析比较，他们得出结论：推进过程中，无论是陆地动物，还是水中生物或空中动物，都遵循相似的能量存储与转换的机理。在产生推进力的同时，也产生可观的侧向力。侧向力对保持稳定性、机动性有作用。产生推进需要肌肉、骨骼、神经、呼吸和循环等系统之间的协调与干预。其中，肌肉在推进中扮演相当多的角色，能起到发动机、刹车、弹簧和支撑等作用。

总之，就像迪金森等六名科学家所说的，生物的任何组织给外部环境（陆地、水、空气）施加作用，外部环境就给生物推进力，道理没有别的，就是牛顿定律。推进力直接来源于压力或摩擦力。

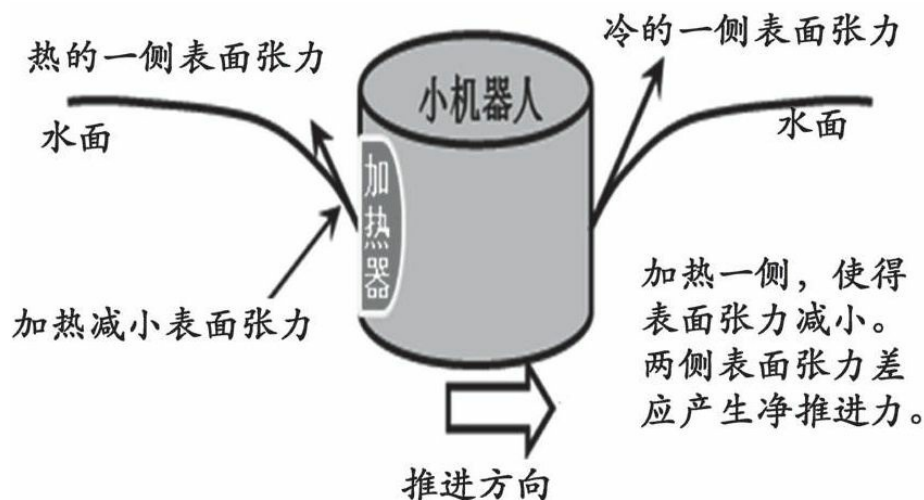


图3.63 表面张力机器人

作用在推进体上的压力（气压、水压或者固体压力）和摩擦力的总和就是推进力。到了这里我们会问，除了气压和水压以及摩擦力，有没有别的直接力可以产生推进。其实，你能想象的所有力，包括电磁力，都可以产生推进。这里举一个表面张力推进的例子。

把一个涂有疏水材料的浮标放在水上，浮标与水面接触线上就有表面张力，但四周的力相互抵消，一般不会因表面张力让浮标自己动起来。使浮标一侧加热，那么水被加热后，当地的表面张力系数会减小，未被加热的一侧的表面张力提供的拉力更大，这样小浮标就可能向冷的一侧运动。如此，就得到了一种可在水面移动的小机器人。你可以自己找找，是否有生物靠表面张力差来产生推进。

刚才说了，推进的本质道理没有别的，就是牛顿定律。然而，下结论也不要那么绝对。罗杰·索耶尔（Roger Shawyer）于1999年提出了一种微波推进系统，称为EmDrive。向一个密封的微波共振腔中注入能量，系统因密封不与环境发生作用，却能产生推进力。该系统显然违背牛顿定律因而一直受质疑。但越来越多的证据表明，该系统确实能产生

推进力。注入一千瓦的微波能量，可以产生数百毫牛的推进力。如果这种推进原理能进一步得到确认可行，那么可能成为星际航行的有效推进手段。

5. 针尖效应 雪花 雪花铃

古人云：草木之花多五出，独雪花六出。可见人们注意到了六角形的雪花。雪花剪纸和雪花图片都是那么美丽，形态令人着迷，以致我们很容易不假思索地认为雪花就是那种长满规则晶针或晶片的美丽形状。可是，我们可能看到的真实雪花又是另外的形状，至少落在地上了的像蓬松的白沙子似的。我们对雪花的印象到底是被艺术家的手笔迷惑了，还是真的因为看到了美轮美奂的雪花形态？雪花飞舞，飘落中有时直线而下，有时在荡秋千似的。落地时悄无声息，落水呢？科学家探测到，雪花落水时，还会发出铃声。

针尖效应与环境对针尖效应的影响

所谓针尖效应，就是针头和钉子很容易渗透到别的物体之中。用到植物上，就是头部越尖的叶子和草，越容易在尖头部位生长，越容易长得长，因为环境中的水分与热量，更容易从尖的部位传入。之所以在这里先提针尖效应和环境对针尖效应的影响，是因为雪花生长受这些效应的影响。

酒瓶兰这种植物的叶很尖，叶子长得很长。羊齿草也是如此。丝瓜藤的叶子有五个瓣，一点都不尖，一点都不细长。为何尖的树叶和草，往往会长得更长？因为它们要从大气中吸收水汽和热量。以热量为例，之所以能吸收热量，是因为有温差。就像暖气片之所以能加热室内空

气，是因为暖气片温度更高。不单单是温差起作用，热量的传递速度更与温度梯度有关。温度梯度是相邻两点之间的温度差除以这两点之间的直线距离。有个定律叫傅里叶热传导定律，指热量传递速率与温度梯度成正比。更热一点的大气，往更冷一点的尖物体传递热量时，越尖就越容易使温度梯度更高，因为同样温差所对应的距离只会由尖细程度反映的距离来决定。因此，大气更容易向尖的头部传递热量。同理，大气更容易向尖的叶子和草的尖头位置传递水分。这就使越尖的地方长得越快，越尖的叶子和草就长得越长，因为除了尖头，其他部位包括侧面都不尖，吸收水分和热量的速率低。

你可以收集一下不同植物叶子的长度与尖顶角的大小，看看是否满足一些能进行表述的反比关系。

环境条件也会有影响。如果大气很干燥，尖的部位和钝的部位都只能依稀地获得水分，谁吸收能力强都不管用了。尖头虽然吸收能力更强，如果环境的水分不够，就无法让其肆意发挥吸收能力。就像你摘苹果的能力很强，一分钟能摘下10个，但如果苹果很少，一分钟也难找到一个，那么你一小时摘到的总的苹果数和那些比你摘苹果能力差的会一样多。正因为如此，干燥的沙漠中，更容易长出仙人球之类的一点也不尖的植物。

类似的例子还很多，现在可以回到雪花了。冰晶的分子呈六角形，角点比边更尖，因此更能从角点抓捕大气中的水雾，于是越是角点越容易生长，长出长长的冰晶芽并不奇怪。类似于上面所说的环境因素影响，如果大气很干燥，那么就没这么多水分任由冰晶芽长出来，而是各个部分长得差不多同样快，于是雪花更容易出现简易的片状结构。如果大气很潮湿，那么吸收能力更强的角点就能得到快速生长的水分，冰晶芽就更容易生长。

有了这些道理想通的知识，就不难理解什么条件下，雪花有什么形态了。

千奇百怪的雪花形态（图3.64、图3.65）

由于冰晶具有六角形分子结构，因此更容易长出六个角。难怪草木之花多五出，独雪花六出。

虽然如此，也不一定是正六角形雪花。有六个角的雪花可能是一个正六边形（片状雪花），也可能是每个角像一把扇子（星星雪花），或者说每个角都是像羊齿草那样的树突（星星树突状或羊齿状雪花）。如前所述，具体形状应与形成雪花的当地温度和湿度等条件有关。

雪花学家肯尼斯·里布李琪（Kenneth Libbrecht）在他的多部专著中，用科学与艺术的眼光近距离观察雪花，试图揭开雪花的秘密世界。站在几何的角度，人们可以把能看到的雪花画成简化的示意图，其类别就有35种。晶芽或枝突的指向、粗细与数目、晶片的形状、大小与数目，都会决定雪花的形态。这么多组合的可能性，恐怕只有掌握了神来之笔的绘画艺术家，才有可能惟妙惟肖地用平面图画再现。

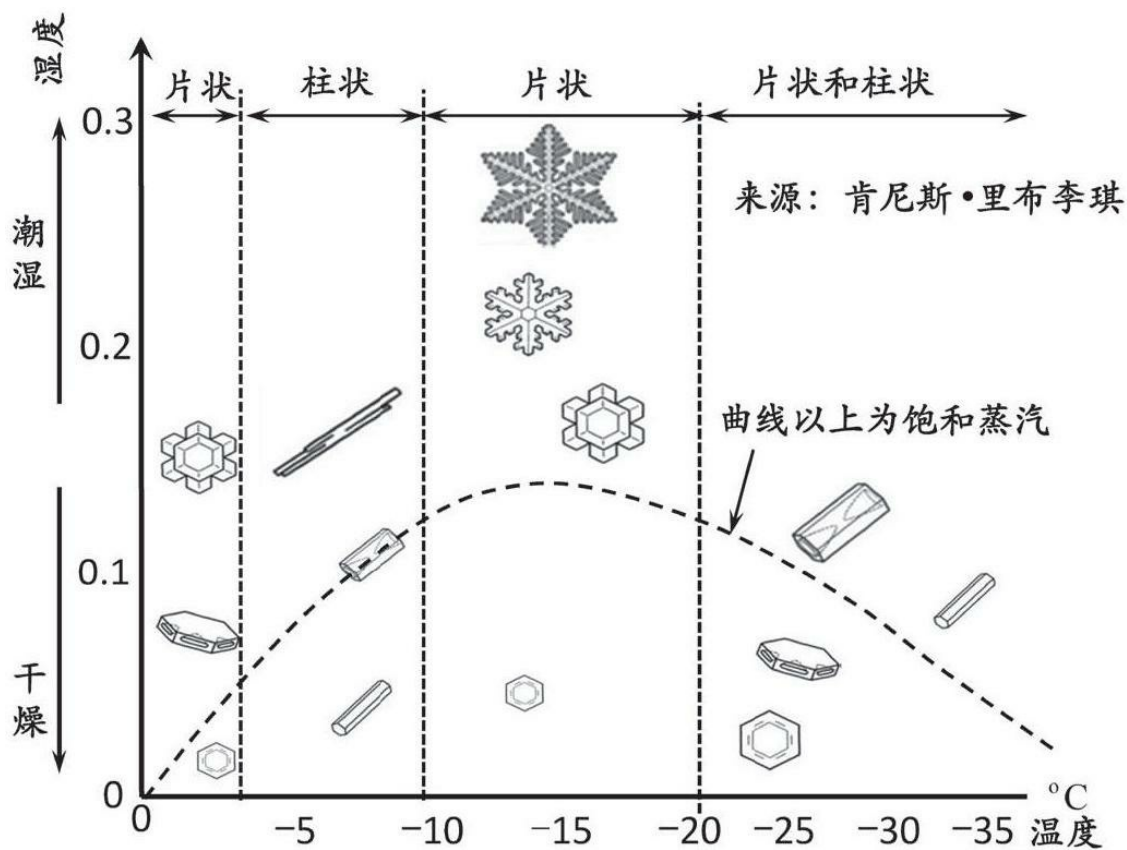


图3.64 不同湿度和温度下的雪花形态

由于针尖效应也受前面介绍的环境效应的影响，因此，按照里布李琪的说法，湿气越大，越容易形成复杂的带晶芽的雪花形状。干燥大气条件下，雪花的形态更简单。温度也有类似影响，因为温度越低越容易普遍结冰，不需要针尖效应帮忙。于是，当温度低于零下22℃时，雪花主要是简单的晶片或晶柱形态。当大气温度较温暖时，雪花会产生更多的分叉（如晶芽）。气温零下15℃左右，主要由湿度和针尖效应决定形状。温度更高以后，可能结冰困难，形成的结构也相对简单。



图3.65 可能的雪花形态简化示意图

2005年，乔恩·尼尔森（Jon Nelson）通过理论分析发现雪花的形态与雪花形成时气象条件的变化有关。如果在雪花形成过程中，气象条件不怎么变化，那么容易形成简单形状。如果雪花在形成过程中，气象条件突然变化，那么来不及均匀地照顾雪花各个部位，而只能在针尖的位置生长，因此气象条件突然变化时，更容易出现长满晶芽的雪花。

人们已经掌握了不同形式的雪花形成的条件，包括大气温度和湿度。因此观察一处的雪花形态，就可以反推出高空的某些气象条件。

对雪花的形态形成规律和各种形态产生的条件的了解，除了对雪花好奇以及可以反过来从雪花形态了解气象外，还可以将雪花冰晶形成的

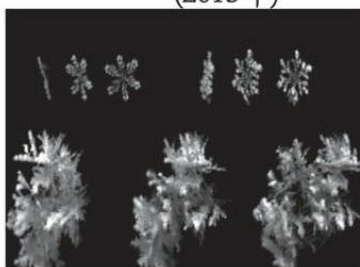
原理应用到结晶学方面。雪花是自然形成的，自然本身的规律往往可以借鉴到技术上。

正如里布李琪所说，雪花是艺术的作品，是大自然的杰作，每一片雪花不同于另外一片，就像每一个人不同于另外一个人。由于雪花的这种多样性，诞生了雪花喜爱者、雪花迷，以及专业或业余的雪花科学家。

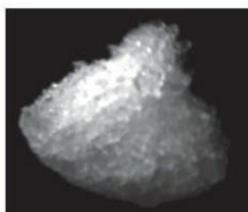
雪花在云端形成时，主要有片晶以及针晶或晶柱两种初始形态。针晶两端尖，像针一样，柱晶的横截面呈六边形（因为冰的分子结构是六边形），柱体长度比六边形直径大很多。片晶呈六角形，比较薄。这些初始形态通过抓捕周围的小水珠，凝结成为雪花。

雪花到底是什么形态？（图3.66）

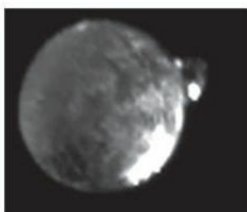
来源：佛勒盖特和加勒特
(2013年)



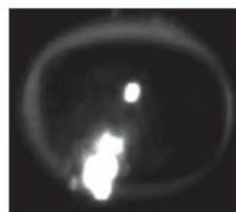
来源：宾利
(19世纪)



雪丸，霰



冰雹



雨滴

图3.66 真实镜头下的雪花、冰雹和雨滴

道理不管多么简单，实际影响因素很多。眼见为实，拍照能给出静态图片，让你看个够。

给雪花拍照可追溯到19世纪。美国人威尔森·宾利在19世纪即拍摄了大量的雪花照片，是他说出了“没有两片雪花是相似的”。他给出的雪花是优美的对称形状。受其影响，艺术家笔下的雪花，均是一些完美的形状。在人们的常识中，有代表性的雪花一定是完美的对称的六边形结晶。

然而，大西洋另一边的德国人古斯塔夫·赫尔曼通过拍照，对雪花形状的表述与美国人不太一样：“每一片雪花的形状都是独特的，许多都是无规则形状，并不完美。”

到底是拍摄技术的问题还是气象条件差异导致了两位雪花爱好者的结论完全不一样？

2015年，犹他州立大学的工程师卡尔·佛勒盖特（Cale Fallgatter）和气象学家蒂姆·加勒特（Tim Garrett）自行研制了可抓拍高空雪花图像的摄像机。他们拍摄了高空飘落的雪花，发现那种完美的扁平六角形雪花很稀少，但还是有六角形雪花。雪花在飘落过程中，除了生长，还会与其他雪花撞在一起，逐步聚合成大而蓬松的雪化团，可能就是我们所说的鹅毛雪。有的雪花被冰冻的小雨滴撞击，形成雪丸（霰）。

雪花的飘落与落水的铃声（图3.67）

长方形纸片、形状相对简单的树叶以及一些带翼带毛的种子在飘落时，就已经够复杂了。雪花的形状如此奇特，如此千变万化，它们飘落时，是如何扰动空气，是如何边下落边生长，空气被扰动的旋涡是如何

与它们共舞……，恐怕很难描述。

你可以想象任何更规则的物体甚至昆虫在空气中的运动方式，都会在雪花中出现。漂移、摇摆、抖动、翻滚、自旋……。

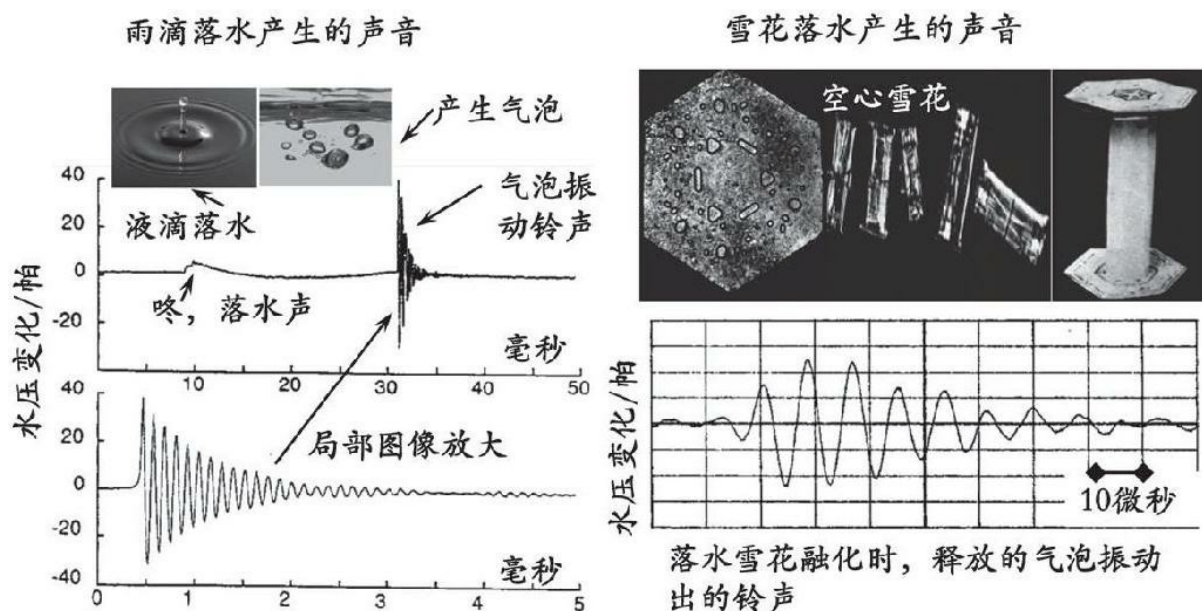
人们用摄像方式测量了雪花飘落的速度，大致是每秒1米左右。我们人走路的速度大概是每秒接近1.5米，因此我们可以一边走路一边追逐漂移的雪花。

雪花飘落时，也在扰动空气，显然会引起声波。雨滴落水时除了咚的一声，紧接着还有铃声，我们在下雨天应该能听出来。那么雪花落水呢？

1999年，美国和英国的四位科学家（克拉姆、庞弗里、罗伊、普罗斯佩雷蒂）用科学仪器倾听了雪花落水的声音，发现雪花落水时，也会发出铃声。这种铃声可以从背景噪声中分辨出来。

雨滴落水时的铃声是这样来的。雨滴落水会打出小气泡。小气泡按固有频率震荡，就发出了铃声。可是雪花太轻，落水时击打不出小气泡。

科学家进一步拨开了雪花的神秘外纱：雪花原来是空心的，是一些空心柱的松散结合体。因此，大多数情况下，雪花的密度只有水的密度的10%。这种蓬松的雪花落水时，不会像水滴那样激起涟漪与气泡，而是突然停在水面上。水的表面张力使水很快蔓延到雪花上，将其融化。于是，雪花中的小气泡聚合在一起在水中释放，如同雨滴落水后形成的气泡，气泡在水中震荡，就会产生铃声。



来源: 气动声学杂志J. Acoust. Soc. Am. 第106卷

图3.67 雨滴和雪花落水产生的声音

琴弦振颤, 可闻曼妙乐曲, 水中小气泡炫舞, 自有响铃之声。既然有铃声, 且测得的声音强度在10分贝量级, 那么我们应该可以听到水面的雪花铃了吧?

可是, 非常令人失望: 这种雪花铃的频率一般高于5万赫兹, 超出了人类2万赫兹的听觉上限。小气泡的振荡频率近似与半径成反比关系, 雪花中小气泡半径非常小, 因此振荡频率非常高。

不知水中的鱼群, 能否听到这种雪花铃?

后记

许多年来，积累了不少与自己专业相关与不相关的知识，因此，总想写一本小型读本，分享给具有各种知识结构的读者群。作为科普读物，我们认为数据和示意图很重要。因此本书包含了大量数据信息和示意图。为了形象和全面，许多示意图需要截取一些公开的图片。有的能找到出处从而尽量在图片上或文中注明（虽然已经是按要求处理过的），也有的无法确定来源；又由于这是一部科普读物，因此除了少量情况外，没有采用专业书籍那样对知识的来源包括摘录的段落进行一一引用，因此只能在这里统一致谢。

本书有不少轻松的内容，如打水漂、蒲公英和棒棒糖等，这些内容以强调趣味性和知识性为主。也有一些深奥的内容，比如演化的普适规律，但因为有其简单、简约和简美的一面，因此即使面对在读的中学生，也觉得有必要分享。鸟和昆虫之类的飞行原理表面上很复杂，但我们尽量用可理解的道理来解释。

绝大部分内容来自于大范围的阅读取材和思考的结果。以数字7的内容为例，阅读的资料甚至涉及一百多年前的文献（反而是近期文献好找）。在写这部分内容时，能找到并参考了的早期文献就有下面4篇[\[1\]](#)。据说，12世纪葡萄牙人纳斯玛尼德斯就解释了数字7是自然世界的神奇数字，但我们没有找到文献。又如雪花，阅读的专业与科普类文献就不下二十篇。如果说要给一个总的文献数目，应该在3000篇以上。

本书涉及的数据信息，确认准确性花了许多时间。某些问题请教过一些专家。例如，玩具陀螺为何不倒是否有更通俗解释，就请教了多名专家。某些取材和解读问题的方式是否恰当或必要，与一些学生讨论过。有的数据还经过演算得到，比如说陀螺不倒时复合离心力的计算就是学生帮助做的。虽然如此，错误和失误在所难免。

本科普书只想给读者分享一些知识和思考解读问题的方法。是否恰当，还需读者评判。如果这部书对读者（包括在校的大中学生）有点阅读价值，那么写这部书的目的就达到了。

[1] 第1篇：Joseph von Hammer-Purgstall（约瑟夫·冯·锤-浦格斯塔），Über die Zahl Sieben (Wiener Jahrbücher der Literatur.. CXXII.182-225. CXXIII.1-54. CXXIV.1-105, 1848年.

第2篇：James Hadley LL.D（杰姆斯·哈德利法律博士），The Number Seven, Essays Philological and Critical (New York: Holt & Williams, 1873), page 325, 1858年.

第3篇：Helena Petrovna Blavatsky（海伦娜·佩特罗弗纳·布拉瓦茨基夫人），The Number Seven, Theosophist, June. 1880年.

第4篇：George A. Miller（乔治·米勒尔），The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, Psychological Review, 63, 81-97, 1956年.

理解科学丛书·张邦宁科普著作

The road
to heaven

MISTY SKY

迷蒙星空

探天之路

张邦宁◎编著

从古至今，宇宙科学与航天科技经历了怎样的过程？

发生过哪些大的、著名事件？

人们是如何来看待和认识这些的？对人类自身又产生了何种影响？

特别是时至今日，

人类对我们身在其中的宇宙本质究竟了解到何种程度？

有哪些重要的研究理论和成果？根据国内外众多专著的论述和历史书籍的记载，
本书会对这些一一予以讲述。

清华大学出版社

理解科学丛书

迷蒙星空：探天之路

张邦宁 编著

清华大学出版社
北 京

张邦宁，研究员，1946年11月出生，1966年毕业于北京清华附中预科，1967年上山下乡到黑龙江绥滨农场，1978年考入哈尔滨工业大学精密仪器系光学专业学习，1982年毕业，进入航天部中国空间技术研究院北京空间机电技术研究所工作。

参加了我国第一代返回型遥感卫星、第一代实时传输型对地遥感卫星光学遥感器的研制工作。完成多项型号、课题的科技报告和设计工作。多次赴卫星发射基地执行任务。

作为光学遥感卫星及光学遥感技术专家，曾担任北京空间机电技术研究所科技委副主任，中国卫星地面辐射试验场专家委员会委员。2009年退休返聘，是目前我国最高分辨率的遥感卫星（高分二号）专家委员会委员。

从古至今，宇宙科学与航天科技经历了怎样的过程？发生过哪些大的、著名事件？人们是如何来看待和认识这些的？对人类自身又产生了何种的影响？特别是时至今日，人类对我们身在其中的宇宙本质究竟了解到何种程度？有哪些重要的研究理论和成果？希望读者通过本书得以认识。宇宙处于永恒地变化、运动中。除此之外，世上不存在绝对之物。任何理论，任何思想，都应该受到宇宙不断变化、运动的检验并加以修正，使之趋于完善。从本书讲述的宇宙、航天研究发展历史中诸多事件充分证明了这点。随着科技的进步，探测水平的提高，使人们对宇宙地理解会更加符合宇宙的真实，并相信：宇宙的规律是可以逐步把握的。顺应宇宙的发展方向，破除各种陈旧之束缚，确立对宇宙真实的认知，应用于周围的世界，使我们的社会、生活和思想趋向科学与和谐。具有宇宙之精神，对真理的自由探索与追求，正是我们共同的宇宙观、理念与信仰。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

迷蒙星空：探天之路 / 张邦宁编著. --北京：清华大学出版社，2015

（理解科学丛书）

ISBN 978-7-302-40729-4

I. ①迷... II. ①张... III. ①宇宙—青少年读物 IV. ①P159-49

中国版本图书馆CIP数据核字（2015）第161919号

责任编辑：朱红莲 王 华

封面设计：

责任校对：刘玉霞

责任印制：

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印 刷 者：

装 订 者：

经 销：全国新华书店

开 本：165mm×240mm

印 张：15.25

字 数：215千字

版 次：2015年6月第1版

印 次：2015年6月第1次印刷

印 数：1～ 000

产品编号：064998-01



绘画：张京

谨以本书献给清华大学附属中学，纪念母校诞生一百周年。

前言

THE MISTY UNIVERSE

THE SKY OF THE ROAD

“宇宙、宇宙学、航天、航天科技”这些词汇已为人们耳熟能详。宇宙大爆炸、黑洞的观测确认；宇宙微波背景辐射和恒星光谱红移的测定；人造卫星，飞船，探月、探火星的飞行器，飞离太阳系、飞向更遥远太空等航天活动，为越来越多的人所知晓，引起各个年龄层人们的好奇与兴趣。那么，从古至今，宇宙科学与航天科技经历了怎样的过程？发生过哪些大的著名事件？人们是如何看待和认识这些的？这些事件对人类自身又产生了何种影响？特别是时至今日，人类对我们身在其中的宇宙本质究竟了解到何种程度，有哪些重要的研究理论和成果？根据国内外众多专著的论述和历史书籍的记载，本书会对这些内容一一予以讲述。

记得多年前，夏日晴朗夜晚，仰望星空，无限深邃，天穹广阔，繁星璀璨，给人以无尽的遐想；看到夜空中流星的划过，人造航天器的移行，感叹人生之短暂，宇宙之浩渺。又经历了多少岁月，随着时代的变迁，科学的进步，技术的发展，思想的解放，使人类有了更深切的感悟：宇宙处于永恒的变化、运动中。除此之外，世上不存在绝对之物。任何理论、任何思想，都应该受到宇宙不断变化、运动的检验并加以修正，使之趋于完善。从本书讲述的宇宙、航天发展历史中的诸多事件充分证明了这点。至于某些利益权贵集团，像西方中世纪教廷，为了维持

宗教统治，墨守成规，残酷镇压那些勇于探索、敢于提出新理论的学者，对他们进行宗教审判、迫害，甚至将他们火刑烧死，使西方社会坠入黑暗世纪。这是很为现代文明社会所不齿的。随着科技的进步，探测水平的提高，人们对宇宙的理解会更加符合宇宙的真实，并相信：宇宙的规律是可以逐步把握的。顺应宇宙的发展方向、破除各种陈旧之束缚、确立对宇宙真实的认知、应用于周围的世界，会使我们的社会、生活和思想趋向科学与和谐。具有宇宙之精神，对真理的自由探索与不断追求，正是我们共同的宇宙观、理念与信仰。

按照宇宙科学和航天科技发展的历史进程，我们划分为五个大的时期来顺序讲述：

远古时代关于宇宙的众多神话传说可称为宇宙学说的启蒙阶段，我国史书上有关的天象记录和天文历法记载是其中亮点。这是第一时期。

从公元前6世纪到公元3世纪，在中国，关于早期宇宙科学的各项内容大体已经完备，一个富有特色的初级体系已经建立起来；而古希腊、罗马的学者在宇宙的本源和结构理论上则出现了激烈争斗，此后西方进入黑暗的中世纪。宇宙学沦入了神学深渊，地心说主宰了宇宙学。这是第二时期。

从公元3世纪到17世纪，在中国，在宇宙观念、仪器制造、历法推算和大地测量等方面取得了许多成就，达到高峰，但后来逐渐衰败下来，经历了从繁荣发展—鼎盛—衰落的过程。而西方，熬过了漫长的黑暗世纪后，16世纪哥白尼倡导的日心说，开始把宇宙学从神学中解放出来；17世纪牛顿开辟了以力学方法研究宇宙学的新途径，诞生了经典宇宙学，他发明了天文望远镜并用于天文观测，取得了众多成果。这是第三时期。

十八九世纪，西方学者创立了星云学说；确立了天体演化学科；对恒星进行了大量观测；把以前只限于太阳系的研究扩大到银河系和河外星系；发明了分光镜；宇宙天文观测的设备、技术和方法取得了大的进展，这些工作为现代宇宙学开拓了道路。中国的宇宙科学受到西学的影响，逐步与其融合。这是第四时期。

20世纪到目前，宇宙科学与航天科技有了极大的发展。量子力学和爱因斯坦相对论的创立；哈勃宇宙膨胀理论和宇宙大爆炸理论的重要证实；河外星系谱线红移和微波背景辐射的发现、证实；中国学者恒星演化模型的建立和对宇宙运动现象的有力诠释；航天理论的奠定和众多航天活动实践；航天遥感、遥测技术和空间飞行技术的发明及应用；展现着宇宙科学和航天科技无穷的魅力。现代宇宙学正在充满活力地发展着。这是第五时期。

需要指出，尽管分为五个时期来讲述，但是各个时期是有密切联系的，会有交织、融汇之处，不能把它们截然分开、孤立看待。时期的年代划分也是粗略的。

经过千百年来一代又一代脑力和体力劳动者不懈的发现、发明、研究、探索和创造，灿烂辉煌的宇宙学和航天大厦已经基本建造起来，但仍有一些不解之谜等待人们去破解。如果通过本书引起诸位兴趣，进而加入探求宇宙和航天奥秘的行列中，并在此过程中陶冶情操，那么笔者会由衷地感到欣慰。

宇宙学问题涉及的学科范围广泛，特别是与天文学紧密关联，而按传统学科划分仅是天文学的一个分支，本书则基本用宇宙学来统一涵盖了。

张邦宁

2015年3月

目 录

[前言THE MISTY UNIVERSE THE SKY OF THE ROAD](#)

[1 启蒙时期，迈出推测宇宙第一步](#)

[1.1 古代中国的宇宙及天文学](#)

[1.1.1 天象观测和记录](#)

[1.1.2 历法制定](#)

[1.1.3 宇宙观念及思想](#)

[1.1.4 神话传说](#)

[1.1.5 小结](#)

[1.2 古埃及、美索不达米亚和古印度的宇宙及天文学](#)

[1.2.1 古埃及](#)

[1.2.2 美索不达米亚](#)

[1.2.3 古印度](#)

[1.2.4 小结](#)

[1.3 古希腊、罗马的宇宙及天文学](#)

[1.3.1 创世神话与传说](#)

[1.3.2 关于宇宙的理论](#)

[1.3.3 天象观测和天文历法](#)

[1.3.4 小结](#)

[参考文献](#)

[2 宇宙及天文学的继续发展](#)

[2.1 中国的宇宙及天文学的进展](#)

[2.1.1 宇宙观念](#)

[2.1.2 宇宙思想认识](#)

[2.1.3 飞天的神话传说、牛郎织女和嫦娥奔月](#)

[2.1.4 天文星图、历法和观测仪器](#)

2.1.5 小结

2.2 埃及、美索不达米亚、印度的宇宙及天文学进展

2.2.1 埃及

2.2.2 美索不达米亚

2.2.3 印度

2.2.4 小结

2.3 希腊、罗马的宇宙及天文学进展

2.3.1 中心火焰说

2.3.2 日心说的先驱

2.3.3 物性论

2.3.4 地心说

2.3.5 历法编制、天象观测、记录和天文测量

2.3.6 小结

参考文献

3 宇宙及天文学挣脱桎梏迈入科学

3.1 中国的宇宙及天文学从鼎盛滑向衰落

3.1.1 天文历法和测量

3.1.2 宇宙观念和思想

3.1.3 世界上的首次航天实践

3.1.4 小结

3.2 西方的宇宙及天文学挣脱桎梏，走入科学轨道

3.2.1 中世纪的宗教神学扼杀宇宙及天文学的发展

3.2.2 复兴的前奏

3.2.3 欧洲及周边阿拉伯地区的宇宙、天文学的复兴

3.2.4 哥白尼日心体系的创立和发展

3.2.5 为确立哥白尼日心说继续斗争

3.2.6 牛顿与经典宇宙学的创立

[3.2.7 万有引力定律的证实](#)

[3.2.8 17世纪的天文望远镜及其观测成就](#)

[3.2.9 小结](#)

[参考文献](#)

[4 天文观测新科技与宇宙、天文学新进展](#)

[4.1 天文观察和测试新科技](#)

[4.1.1 天文望远镜技术快速发展](#)

[4.1.2 分光学、光度学、光谱学和照相术应用于宇宙、天文领域](#)

[4.1.3 天体距离的测量](#)

[4.2 经典宇宙理论的创立与发展](#)

[4.2.1 经典宇宙理论的力学和数学基础](#)

[4.2.2 太阳系起源说及康德和拉普拉斯的星云说](#)

[4.2.3 银河系概念的初步确立](#)

[4.2.4 恒星天文学之父——英国天文学家威廉·赫歇尔](#)

[4.2.5 儒勒·凡尔纳的航天科幻小说《从地球到月球》和《环绕月球》](#)

[4.2.6 小结](#)

[参考文献](#)

[5 现代宇宙、天文学理论与宇宙实验观测证实](#)

[5.1 银河系之争及河外星系的确认](#)

[5.1.1 天体距离的观测技术和分析方法](#)

[5.1.2 银河系之争](#)

[5.1.3 河外星系的确认](#)

[5.2 现代宇宙学理论和观测证实的发展](#)

[5.2.1 宇宙膨胀说的提出](#)

[5.2.2 赫罗图和恒星坍塌理论](#)

[5.2.3 相对论与量子力学的发展](#)

[5.2.4 宇宙大爆炸理论](#)

[5.2.5 黑洞理论](#)

[5.3 天体观测的进展](#)

[5.3.1 射电天文学的诞生](#)

[5.3.2 银河系结构的射电探测](#)

[5.3.3 20世纪60年代的四大天文发现](#)

[5.3.4 射电天文学的新进展](#)

[5.4 宇航时代到来](#)

[5.4.1 空间探测手段的发展](#)

[5.4.2 开展空间探测](#)

[5.4.3 载人航天](#)

[5.5 小结](#)

[参考文献](#)

[返回总目录](#)

1 启蒙时期，迈出推测宇宙第一步

远古时代，人类的祖先在日常生活和从事农牧业生产过程中，逐渐意识到日月运行、昼夜交替、寒来暑往这些天象变化对他们影响很大，与他们密切相关，并且变化是有一定规律的。人们通过对天文现象的长期观察、忠实记录，形成了早期的信息积累；并结合实际需要，制定出了历法。天象的观察、记录和历法的订立构成天文学的第1章，也是认识宇宙的开端。与此同时，壮丽的、有规律的、有时却变幻多端的天象也引起了人们的赞叹、惊恐、信服和崇敬，从而产生了对自然力的崇拜。占星术受到统治阶层的重视，他们想要以此来了解上天。智者在久久思索后，作出对宇宙的推测，努力解开各种疑团。在史前考古和古岩画中，发现了有类似飞行器和日月星辰的图形，由此可以看到人类思想的活跃。由于实证极少，这里仅点到为止。经过上述因素综合作用，早期的宇宙学说出现，同时也有了远古的神话和宗教。

1.1 古代中国的宇宙及天文学

1.1.1 天象观测和记录

据《尚书》记载，早在尧舜时代，“乃命羲和，钦若昊天，历象日月星辰，敬授人时”。即按照《尚书》所言，在四千多年前的尧舜时代就已经设立专门的天文官职，任命羲和来观天象，记录日月星辰之变动情况，使人们把握农时季节。

在古代，观测、记录并解释天象，是统治者的特权，是垄断性的，“绝地通天”，普通百姓不得参与。统治者以此表示自己是受命于天，是能够与上天沟通的，是得到上天保护的，自己的一切行动都是符合天意的，百姓必须服从。统治者每年都要举行几次大典，祭祀上天、昭告四方自己是正统，以期继续受到上天保佑。如果观测到凶兆，统治者要想方设法来化解，严重的要下“罪己诏”来检讨。每当军事行动之前，必得观天象、占卜吉凶，期望顺应天意，获取胜利。不过常见的是为农事，以确认季节时令。

《尚书·尧典》，“期三百有六旬有六日，以闰月定四时成岁”。当时确定平年12朔，若干年后加闰月，年平均长度366天。

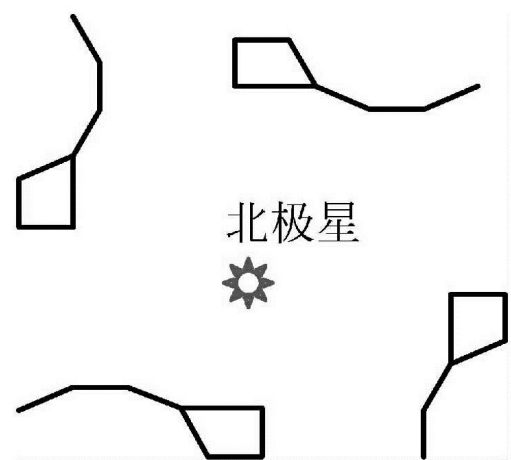
通过观测“昏中星”判断季节。

夏朝后期（公元前18世纪—前16世纪），还观测“旦中星”和北斗星斗柄的指向变化，以便更准确地掌握节令。“斗柄东指，天下皆春；斗柄南指，天下皆夏；斗柄西指，天下皆秋；斗柄北指，天下皆冬。”

商代（公元前16世纪—前11世纪）后期盘庚定都于殷，有大量甲骨

记载各种天象：日食、月食、新星和干支纪日等。

西周时期（公元前11世纪—前8世纪）。把黄道和赤道区域划分为二十八宿，后来进而分为东、南、西、北四宫，并与“四象”相配：



东宫苍龙：角、亢、氐、房、心、尾、箕

南宫朱雀：井、鬼、柳、星、张、翼、轸

西宫白虎：奎、娄、胃、昂、毕、

北宫玄武：斗、牛、女、虚、危、室、壁

开始用仪器进行天文观测，已发明圭表。西周初期周公在阳城（今河南省登封市告城镇）设立测景台。

关于日月星辰的记载：

1. 日食和月食

殷商甲骨文有日月食记载，留存的甲骨卜辞中，至少有三条被认为描述了日食现象，有五条被证实是可靠的月食记录。最早的记载为《尚书·胤征》记录夏仲康元年（约公元前2000年）“乃季秋月朔，辰弗集于房”。

日食分为“日全食”、“日环食”和“日偏食”3种。除了食分很少的日偏

食不易引起人们注意外，日食总体来讲是一种非常显著的天象变化。月食分为“月全食”和“月偏食”2种，其中月偏食不易引人注意。中国古代有世界上最完整的日食及月食记录。

《诗·小雅·十月之交》中记载了一次日食和一次月食，大致确定发生在公元前8世纪前期。

《逸周书·小开解》中记载：“维三十有五祀，五念曰：正月丙子，拜里食无时。”即文王三十五年正月丙子发生了一次月食。据认为这次月食也是在商纣王三十八年，故将这次月食作为联系殷、周年代的确证。

《春秋》中记载的日食和月食大都可考。《春秋》以后的日食和月食都有史可查。中国古代史籍中关于日食的记载共计1600多项，月食为1100多项。这是中国古代宇宙、天文学遗产的宝贵组成部分。

2. 太阳黑子

最早记录为汉成帝河平元年（公元前28年）：“汉成帝河平元年三月乙未，日出黄，有黑气大如钱，居中央。”

从此以后直至清代，共有200多次黑子记录。

3. 彗星

最早记录为殷末（公元前11世纪）武王伐纣时所见的彗星，《淮南子·兵略训》：“武王伐纣，……彗星出而授殷人其柄。”

从此以后直至清代，彗星记录共有2000余次。

关于哈雷彗星的最早记录是春秋鲁文公十四年（公元前613年）。自秦始皇七年（公元前240年）至公元1910年，共29次回归，每次都有记录。并且详细记录了数次周期彗星的回归。

4. 新星和超星

从公元前1300年至公元1700年中国记录新星或新星爆发68次。

关于公元1054年发现的超新星，《宋会要》记载：“至和元年五月，晨出东方，守天关。昼见如太白，芒角四出，色赤白，凡见二十三日。”1054年7月4日爆发的超新星在金牛座附近，一直到1056年4月6日才肉眼不可见。

5. 流星、流星雨和陨石

从西周至清代共5000余条这类关于流星和流星雨的记录，约500次关于陨石降落。

最早的流星雨记载在《春秋》：“鲁庄公七年四月辛卯（公元前687年3月23日）夜，恒星不见，夜中星陨如雨。”

最早陨石降落记载在《春秋》里：“僖公十有六年春正月戊申朔（公元前645年12月24日），陨石于宋（今河南商丘），五。”

6. 掩星

掩星大致可分为4种情况：月掩恒星、月掩行星、行星掩恒星和行星互掩等，其中月掩恒星最为常见，而月掩行星最为古天文学家所重视，所称掩星是指月掩行星。《中国古代天象记录总集》记载历史上的月掩行星共有200多次。这些记录对验证现代宇宙学的某些理论，如地

球自转速度的长期变化是有意义的。

7. 星表和星图

战国时代魏国的石申和齐国的甘德各编星表，后人合称《甘石星经》。其中记载了121颗恒星的坐标，这是已知的我国最早的星表。

1.1.2 历法制定

据《汉书·艺文志》中记载：“黄帝五家历三十三卷，颛顼历二十一卷，夏殷周鲁历十四卷。”这就是古六历。结合考古发现，4000多年前就已经有了历法。

春秋时期的历法：一年分为十二个月；正月、二月和三月为春季；每3个月为一季，春夏秋冬，以此类推。规定以日月合朔那天为初一，称为朔日；每个月最后一天（二十九日或三十日）称为晦日。因为十二个月共354天或355天，短于一个回归年（当时认为是365.2423天一个回归年）。为了使月份和冷暖季节大致相对应，所以每隔2年或3年就必须插入一个闰年，这被称作阴阳历。

中国古代历法的主要特点：

（1）用干支纪日、岁星纪年和干支纪年法。

从商代已实行干支纪日法。从春秋鲁隐公元年（公元前722年）二月己巳日起日干支从未间断。

（2）用二十四节气。

（3）非常重视朔的推算。

（4）涉及内容十分广泛。

古代中国天象观测和历法的发展是中国天文学史的一条主线。

1.1.3 宇宙观念及思想

盖天说是中国古代最早的一种宇宙学说，这一学说大约源于公元前11世纪的殷末周初，有“旧盖天说”“与新盖天说”之分。旧盖天说认为“天圆如张盖，地方如棋局”，穹隆状的天幕盖在正方形的平直大地上。新盖天说则认为“天似盖笠，地法覆盎”，“以《周髀算经》为基本纲领性文献，提出了自成体系的定量化天地结构。”这一学说流传至今，被当作是中国古代天文学说的鼻祖。但是，它的错误是明显的，经不起推敲的，当时就有学者对其提出质疑。然而它是最早的，是历史逐步演变发展过程，故作一简单介绍。

《晋书·天文志》中记载有《周髀算经》中的观点：“其言天似盖笠，地法覆盎，天地各中高外下。北极之下为天地之中，其地最高，而滂沲四环，三光隐映，以为昼夜。天中高于外衡冬至日之所在六万里。北极下地高于外衡下地亦六万里，外衡高于北极下地二万里。天地隆高相从，日去地恒八万里。”

大概意思是：天是一个穹形，地也是一个穹形，二者如同同心半球一般。两个穹球的间距是八万里。北极是天穹的最高中央，日月星辰周转不息，成为白天黑夜。日月星辰的出没，不过是距离远近使然。

约2500年前的春秋战国时期，思想家老子在其著作《老子》（也称《道德经》）中就提出了以“道”（即规律）解释宇宙万物的演变，认为“道生一，一生二，二生三，三生万物”，“道”乃“夫莫之命而常自

然”，因而“人法地，地法天，天法道，道法自然”的自发的唯物主义观点。《老子》一书中还包含有大量朴素辩证法观点，如认为一切事物均具有正反两面，“反者道之动”，并能由对立而转化。“天地无人推而自行，日月无人燃而自明，星辰无人列而自序，禽兽无人造而自生，此乃自然为之也，何劳人为乎？”



“天长地久。天地所以能长且久者，以其不自生，故能长生。是以圣人后其身而身先；外其身而身存。非以其无私邪？故能成其私。”

老子

“上善若水。水善利万物而不争，处众人之所恶，故几于道。”

“居善地，心善渊，与善仁，言善信，政善治，事善能，动善时。夫唯不争，故无尤。”

“持而盈之，不如其已；揣而锐之，不可长保。金玉满堂，莫之能守；富贵而骄，自遗其咎。功遂身退，天之道也。”

“三十辐，共一毂，当其无，有车之用。埴埴以为器，当其无，有器之用。凿户牖以为室，当其无，有室之用。故有之以为利，无之以为用。”

“五色令人目盲；五音令人耳聋；五味令人口爽；驰骋畋猎，令人心发狂；难得之货，令人行妨。是以圣人为腹不为目，故去彼取此。”

“宠辱若惊，贵大患若身。”

“大道废，有仁义；智慧出，有大伪；六亲不和，有孝慈；国家昏乱，有忠臣。”

“绝圣弃智，民利百倍；绝仁弃义，民复孝慈；绝巧弃利，盗贼无有。此三者以为文，不足。故令有所属：见素抱朴，少思寡欲，绝学无忧。”

“人之所畏，不可不畏。”

“俗人昭昭，我独昏昏。俗人察察，我独闷闷。”

“曲则全，枉则直，洼则盈，敝则新，少则多，多则惑。是以圣人抱一为天下式。不自见，故明；不自是，故彰；不自伐，故有功；不自矜，故长。夫唯不争，故天下莫能与之争。”

“希言自然。故飘风不终朝，骤雨不终日。”

“企者不立；跨者不行；自见者不明；自是者不彰；自伐者无功；自矜者不长。”

“有物混成，先天地生。寂兮寥兮，独立而不改，周行而不殆，可以为天地母。吾不知其名，强字之曰道，强为之名曰大。大曰逝，逝曰远，远曰反。”

“故道大，天大，地大，人亦大。域中有四大，而人居其一焉。”

这些书中的经典名言，体现着老子远远超出那个时代的宇宙思想，至今仍能给人们以启迪、值得我们认真研究、继承。

老子，姓李名耳，字聃，华夏族，楚国苦县厉乡曲仁里（今河南省鹿邑县太清宫镇）人，大约生活于公元前571年至公元前471年之间，是我国古代伟大的哲学家、思想家，是道家学派创始人，世界百位历史名人之一。老子的著作、思想早已成为世界文化遗产中的宝贵财富。欧洲从19世纪初就开始了对《道德经》的研究，到20世纪的四五十年代，欧洲共有60多种《道德经》译文。在当下的德国、法国、英国、美国、日本等发达国家相继兴起了“老子热”，《老子》一书在这些国家被一版再版。20世纪80年代，据联合国教科文组织统计，在世界文化名著中，译成外国文字出版发行量最大的是《圣经》，第二位就是《道德经》。

中国古代在老子的《道德经》之前，还有一部体现最早的宇宙思想的奇书——《易经》（即《周易》）。1973年初，长沙马王堆汉墓出土了帛书《周易》，后又在阜阳双古堆汉墓出土了汉简《周易》。还发现了楚竹书《周易》。相传《周易》为周文王（公元前1152—前1056年）所作，其后有所修改。《史记·太史公自序》中说：“昔西伯（即周文王）拘羑里，演《周易》。”“大抵贤圣发愤之所为作也。”在艰险困苦的条件下，发奋编写出了传世奇书《易经》。



八卦图

《童子问易》指出：《易经》持有的是生生不息的宇宙观，是本体论的生命哲学。它既讲世界观，又讲方法论。其认为世界是“有”（太极）；其方法论包括辩证法和一分为二（“分阴”、“分阳”）。还

有“复”观的方法，要求当见“天地之心”。天地之心不是所谓的“道心”，而是德心；有“穷则思变、变以从道”的方法；“极深研几、见机而作”的方法；“钩深致远”的方法；物相杂、“杂而不越”的方法；“善持盈”的方法；“进德修业”等方法。从《易经》中体悟的天地、人事现象背后的隐约规则，在无穷变化中有一个不变的太极，由此而生两仪，两仪再变生四象，四象演化为八卦，极致到无穷。《周易》是一部古老而又灿烂的文化瑰宝，古人用它来预测未来、决策国家大事、反映当前现象，上测天，下测地，中测人事。然而《周易》占测只属其中的一个功能，其实《周易》囊括了天文、地理、军事、科学、文学、农学等丰富的知识内容，许多宇宙思维、理念还有待于人们去进一步挖掘、破解。

春秋战国时代另一位学者墨子则具有朴素的宇宙思想，在许多方面都很有成就。墨子（公元前468年—前376年），名翟，汉族，宋国国都（今河南商丘）人，是战国时期著名的思想家、教育家、科学家、军事家，是墨家学派的创始人及主要代表人物。

墨子认为，宇宙是一个连续的整体，个体或局部都是由这个统一的整体分出来的，都是这个统一整体的组成部分。换句话说，也就是说整体包含着个体，整体又是由个体所构成，整体与个体之间有着必然的有机联系。从这一连续的宇宙观出发，墨子进而建立了关于时空的理论。他把时间定名为“久”，把空间定名为“宇”，并给出了“久”和“宇”的定义，即“久”为包括古今旦暮的一切时



间，“宇”为包括东西中南北的一切空间，时间和空间都是连续不间断的。墨子

在给出了时空的定义之后，墨子又进一步论述了时空有限还是无限的问题。他认为，时空既是有穷的，又是无穷的。对于整体来说，时空是无穷的；而对于部分来说，时空则是有穷的。他还指出，连续的时空是由时空元所组成。他把时空元定义为“始”和“端”，“始”是时间中不可再分割的最小单位，“端”是空间中不可再分割的最小单位。这样就形成了时空是连续无穷的，这连续无穷的时空又是由最小的单元所构成，在无穷中包含着有穷，在连续中包含着不连续的时空理论。

在时空理论的基础上，墨子建立了自己的运动论。他把时间、空间和物体运动统一起来，联系在一起。他认为，在连续的、统一的宇宙中，物体的运动表现为在时间中的先后差异和在空间中的位置迁移。没有时间先后和位置远近的变化，也就无所谓运动，离开时空的单纯运动是不存在的。

对于物质的本原和属性问题，墨子也有精辟的阐述。在先秦诸子中，老子最早提出了物质的本原是“有生于无”，“天下万物生于有，有生于无”。墨子首先起来反对老子的这一思想，提出了万物始于“有”的主张。他指出，“无”有两种，一种是过去有过而如今没有了，如某种灭绝的飞禽，这不能因其已不存在而否定其曾为“有”；一种是过去就从来没有过的事物，如天塌陷的事，这是本来就不存在的“无”。本来就不存在的“无”不会生“有”，本来存在后来不存在的更不是“有”生于“无”。

由此可见，“有”是客观存在的。接着，墨子进而阐发了关于物质属性的问题。他认为，如果没有石头，就不会知道石头的坚硬和颜色，没有日和火，就不会知道热。也就是说，属性不会离开物质客体而存在，

属性是物质客体的客观反映。人之所以能够感知物质的属性，是由于有物质客体的客观存在。

墨子是中国历史上第一个从理性高度对待数学问题的科学家，他给出了一系列数学概念的命题和定义，这些命题和定义都具有高度的抽象性和严密性。

墨子所给出的数学概念主要有：

关于“倍”的定义。墨子说：“倍，为二也。”（《墨经上》）亦即原数加一次，或原数乘以二称为“倍”。如二尺为一尺的“倍”。关于“平”的定义。墨子说：“平，同高也。”（《墨经上》）也就是同样的高度称为“平”。这与欧几里得几何学定理“平行线间的公垂线相等”意思相同。

关于“同长”的定义。墨子说：“同长，以正相尽也。”（《墨经上》）也就是说两个物体的长度相互比较，正好一一对应，完全相等，称为“同长”。

关于“中”的定义。墨子说：“中，同长也。”（《墨经上》）这里的“中”指物体的对称中心，也就是物体的中心为与物体表面距离都相等的点。

关于“圜”的定义。墨子说：“圜，一中同长也。”（《墨经上》）这里的“圜”即为圆。墨子指出圆可用圆规画出，也可用圆规进行检验。圆规在墨子之前早已得到广泛的应用，但给予圆精确的定义，则是墨子的贡献。墨子关于圆的定义与欧几里得几何学中圆的定义完全一致。

关于正方形的定义。墨子说，四个角都为直角，四条边长度相等的四边形即为正方形，正方形可用直角曲尺“矩”来画图 and 检验。这与欧几

里得几何学中的正方形定义也是一致的。

关于直线的定义。墨子说，三点共线即为直线。三点共线为直线的定义，在后世测量物体的高度和距离方面得到广泛的应用。三国时期的刘徽在测量学专著《海岛算经》中，就是应用三点共线来测高和测远的。汉以后弩机上的瞄准器“望山”也是据此发明的。

此外，墨子还对十进位值制进行了论述。中国早在商代就已经比较普遍地应用了十进制记数法，墨子则是对位值制概念进行总结和阐述的第一个科学家。他明确指出，在不同位数上的数码，其数值不同。例如，在相同的数位上，1小于5，而在不同的数位上，1可多于5。这是因为在同一数位上（个位、十位、百位、千位……），5包含了1；而当1处于较高的数位上时，则反过来1包含了5。十进制的发明，是中国对于世界文明的一个重大贡献。正如《中国科学技术史》的数学卷中所说：“商代的数字系统比古巴比伦和古埃及同一时代的体系更为先进、更为科学的”，“如果没有这种十进位制，就几乎不可能出现我们现在这个统一化的世界了”。

墨子关于物理学的研究涉及力学、光学、声学等分支，给出了不少物理学概念的定义，并有不少重大的发现，总结出了一些重要的物理学定理。

首先，墨子给出了力的定义，说：“力，刑（形）之所以奋也。”（《墨经上》）也就是说，力是使物体运动的原因，即，使物体运动的作用叫做力。对此，他举例予以说明，例如将重物由下向上举，有力的作用就能做到。同时，墨子指出物体在受力之时，也产生反作用力。例如，两个质量相当的物体碰撞后，各自就会朝相反的方向运动。如果两个物体的质量相差较大，碰撞后质量大的物体虽不会动，但反作

用力还是存在。

接着，墨子又给出了“动”与“止”的定义。他认为“动”是力推送的缘故，“止”则是物体经过一定时间后运动状态的结束。墨子虽没有明确指出运动状态的结束是因为存在着阻力的缘故，但他已意识到在外力消失后，物体的运动状态是不可能永远存在下去的。

关于杠杆定理，墨子也作出了精辟的表述。他指出，称重物时秤杆之所以会平衡，原因是“本”短“标”长。用现代的科学语言来说，“本”即为阻力臂，“标”即为动力臂，写成为力学公式就是“动力×动力臂”（“标”）=阻力×阻力臂（“本”）。此外，墨子还对杠杆、斜面、重心、滚动摩擦等力学问题进行了一系列的研究，这里就不一一赘述。在光学史上，墨子是第一个进行光学实验，并对几何光学进行系统研究的科学家。如果说墨子奠定了几何光学的基础，也不为过分，至少在中国是这样。正如《中国科学技术史》中所说，墨子关于光学的研究，“比我们所知的希腊的为早”，“印度亦不能比拟”。

墨子首先探讨了光与影的关系，他细致地观察了运动物体影像的变化规律，提出了“景不徙”的命题。也就是说，运动着的物体从表观看，它的影也是随着物体在运动着，其实这是一种错觉。因为当运动着的物体位置移动后，它前一瞬间所形成的影像已经消失，其位移后所形成的影像已是新形成的，而不是原有的影像运动到新的位置。如果原有的影像不消失，那它就会永远存在于原有的位置，这实际上是不可能的。因此，所看到的影像的运动，只是新旧影像随着物体运动而连续不间断地生灭交替所形成的，并不是影像自身在运动。墨子的这一命题，后来为名家所继承，并由此提出了“飞鸟之影未尝动”的命题。

随之，墨子又探讨了物体的本影和副影的问题。他指出，光源如果

不是点光源，由于从各点发射的光线产生重复照射，物体就会产生本影和副影；如果光源是点光源，则只有本影出现。

接着，墨子又进行了小孔成像的实验。他明确指出，光是直线传播的，物体通过小孔所形成的像是倒像。这是因为光线经过物体再穿过小孔时，由于光的直线传播，物体上方成像于下，物体下部成像于上，故所成的像为倒像。他还探讨了影像的大小与物体的斜正、光源的远近的关系，指出物斜或光源远则影长细，物正或光源近则影短粗。如果是反射光，则影形成于物与光源之间。

特别可贵的是，墨子对平面镜、凹面镜、凸面镜等进行了相当系统的研究，得出了几何光学的一系列基本原理。他指出，平面镜所形成的是大小相同、远近对称的像，但却左右倒换。如果是2个或多个平面镜相向而照射，则会出现重复反射，形成无数的像。凹面镜的成像是在“中”之内形成正像，距“中”远所成像大，距“中”近所成的像小，在“中”处则像与物一样大；在“中”之外，形成的是倒像，近“中”像大，远“中”像小。凸面镜只形成正像，近镜像大，远镜像小。这里的“中”为球面镜的球心，墨子虽尚未能区分球心与焦点的差别，把球心与焦点混淆在一起，但其结论与近现代球面镜成像原理还是基本相符的。

墨子还对声音的传播进行过研究，发现井和罍（小口大肚的瓶子）有放大声音的作用，并加以巧妙地利用。他曾教导学生说，在守城时，为了预防敌人挖地道攻城，每隔三十尺挖一井。置大罍于井中，罍口绷上薄牛皮，让听力好的人伏在罍上进行侦听，以监知敌方是否在挖地道，地道挖于何方，从而做好御敌的准备（原文是：令陶者为罍，容四十斗以上，……置井中，使聪耳者伏罍而听之，审知穴之所在，凿内迎之）。尽管当时墨子还不可能明白声音共振的机理，但这个防敌方法却蕴含着丰富的科学内涵。

墨子是一个精通机械制造的大家，在止楚攻宋时与公输般进行的攻防演练中，已充分地体现了他在这方面的才能和造诣。他曾花费了3年的时间，精心研制出一种能够飞行的木鸟（风筝），成为我国古代风筝的创始人。他又是一个制造车辆的能手，可以在不到一日的时间内造出载重30石的车子。他所造的车子运行迅速又省力，且经久耐用，为当时的人们所赞赏。

值得指出的是，墨子几乎谙熟当时各种兵器、机械和工程建筑的制造技术，并有不少创造。在《墨子》一书中的“备城门”、“备水”、“备穴”、“备蛾”、“迎敌祠”、“杂守”等篇中，他详细地介绍和阐述了城门的悬门结构，城门和城内外各种防御设施的构造，弩、桔槔和各种攻守器械的制造工艺，以及水道和地道的构筑技术。他所论及的这些器械和设施，对后世的军事活动有着很大的影响。

墨子的哲学建树以认识论和逻辑学最为突出，其贡献是先秦其他诸子所无法比拟的。

墨子认为，人的知识来源可分为三个方面，即闻知、说知和亲知。他把闻知又分为传闻和亲闻两种，但不管是传闻或亲闻，在墨子看来都不应当是简单地接受，而必须消化并融会贯通，使之成为自己的知识。因此，他强调要“循所闻而得其义”，即在听闻、承受之后，加以思索、考察，以别人的知识作为基础，进而继承和发扬。

墨子所说的“说知”，包含有推论、考察的意思，指由推论而得到的知识。他特别强调“闻所不知若已知，则两知之”，即由已知的知识去推知未知的知识。如已知火是热的，推知所有的火都是热的；圆可用圆规画出，推知所有的圆都可用圆规度量。由此可见，墨子的闻知和说知不是消极简单地承受，而是蕴含着积极的进取精神。

除闻知和说知外，墨子非常重视亲知，这也是墨子与先秦其他诸子的一个重大不同之处。墨子所说的亲知，乃是自身亲历所得到的知识。他把亲知的过程分为“虑”、“接”、“明”三个步骤。“虑”是人的认识能力求知的状态，即生心动念之始，以心趣境，有所求索。但仅仅思虑却未必能得到知识，譬如张眼睨视外物，未必能认识到外物的真相。因而要“接”知，让眼、耳、鼻、舌、身等感觉器官去与外物相接触，以感知外物的外部性质和形状。而“接”知得到的仍然是很不完全的知识，它所得到的只能是事物的表观知识。有些事物，如时间，是感官所不能感受到的。因此，人由感官得到的知识还是初步的，不完全的，还必须把得到的知识加以综合、整理、分析和推论，方能达到“明”知的境界。总之，墨子把知识来源的三个方面有机地联系在一起，在认识论领域中独树一帜。

墨子又是中国逻辑学的奠基者。他称逻辑学为“辩”学，把其视之为“别同异，明是非”的思维法则。他认为，人们运用思维认识现实，作出的判断无非是“同”或“异”，“是”或“非”。为此，首先就必须建立判别同异、是非的法则，以之作为衡量、判断的标准，合者为“是”，不合者为“非”。这种判断是“不可两不可”的，人们运用思维以认识事物，对同一事物作出的判断，或为“是”，或为“非”，二者必居其一，没有第三种可能存在，不可能二者都为“是”，或二者都为“非”，也不可能既“是”又“非”，或既“非”又“是”。用现代的逻辑学名词来说，这就是排中律和矛盾律。

由这一思维法则出发，墨子进而建立了一系列的思维方法。他把思维的基础方法概括为“摹略万物之然，论求群言之比。以名举实，以辞抒意，以说出故。以类取，以类予”（“小取”）。也就是说，思维的目的在于要探求客观事物间的必然联系，以及探求反映这种必然联系的形

式，并用“名”（概念）、“辞”（判断）、“说”（推理）表达出来。“以类取，以类予”，相当于现代逻辑学的类比，是一种重要的推理方法。此外，墨子还总结出了假言、直言、选言、演绎、归纳等多种推理方法，从而使墨子的辩学形成一个有条不紊、系统分明的体系，在古代世界中别树一帜。墨子是中国古代逻辑思想的重要开拓者之一。墨辩和古代印度的因明学、古希腊的逻辑学并称世界三大逻辑学。他比较自觉、大量地运用了逻辑推论的方法，以建立或论证自己的政治、伦理思想。他还在中国逻辑史上第一次提出了“辩”、“类”、“故”等逻辑概念，并要求将“辩”作为一种专门知识来学习。墨子的“辩”虽然统指辩论技术，但却是建立在“知类”（事物之类）“明故”（根据、理由）基础上的，因而属于逻辑类推或论证的范畴。墨子所说的“三表”既是言谈的思想标准，也包含有推理论证的因素。墨子还善于运用类推的方法揭露论敌的自相矛盾。由于墨子的倡导和启蒙，墨家养成了重逻辑的传统，并在建立了第一个中国古代逻辑学的体系。

在墨家整个思想体系中，军事思想占有重要位置。《墨子》的军事思想是处于弱者地位的自卫学说，其主要内容有二：一是非攻，反对攻伐掠夺的不义之战；二是救守，支持防守诛讨的正义之战。

（1）非攻：反对攻伐掠夺的不义之战

墨子认为，当时进行的战争均属掠夺性非正义战争，在《非攻》诸篇中，反复申诉非攻之大义，认为战争是凶事。他说，古者万国，绝大多数在攻战中消亡殆尽，只有极少数国家幸存。这就好比医生医了上万人，仅仅有几人痊愈，这个医生不配称之为良医，战争同样不是治病良方。历史上好战而亡的统治者不可胜数。这无异于给那些企图通过攻战来开疆拓土吞并天下的人以当头棒喝。所以墨子主张，以德义服天下，以兼爱来消弭祸乱。在墨子眼里，兼爱可以止攻，可以去乱。兼爱是非

攻的伦理道德基础，非攻是兼爱的必然结果。

墨子主张非攻，是特指反对当时的“大则攻小也，强则侮弱也，众则贼寡也，诈则欺愚也，贵则傲贱也，富则骄贫也”的掠夺性战争。墨子以是否兼爱为准绳，把战争严格区分为“诛”（诛无道）和“攻”（攻无罪），即正义与非正义两类。“兼爱天下之百姓”的战争，如禹攻三苗、商汤伐桀、武王伐纣，是上（符合）天之利、中鬼之利、下人之利的，因而有天命指示，有鬼神的帮助，是正义战争。反之，大攻下、强凌弱、众暴寡，“兼恶天下之百姓”的战争，是非正义的。墨子坚决无情地揭发了当时战争给人民带来的沉重无尽的灾难。

（2）救守：支持防守诛讨的正义之战

墨子“惟非攻，是以讲求备御之法”，从“非攻”出发，《墨子》论述了作为弱小国家如何积极防御的问题。墨子深知，光讲道理，大国君主是不会放弃战争的，因而主张“深谋备御”，以积极防御制止以大攻小的侵略战争。

中国古代战争最著名的守城战术典籍是墨家的《墨子》，《墨子》十五卷，其中第十四、十五卷全篇介绍了守城的装备、战术、要点，共二十篇，存十一篇，为《备城门》、《备高临》、《备梯》、《备水》、《备突》、《备穴》、《备蛾傅（即蚁伏，指步兵强行登城）》、《迎敌祠》、《旗帜》、《号令》和《集守》。《墨子》中的守城战术极其丰富，仅存的十一篇就几乎涵盖了所有冷兵器时代的攻城术。

综上所述，可以看到墨子的科学造诣之深、成就之大，在中国古代杰出科学家的行列中堪称佼佼者。但遗憾的是，墨子在科技领域中的理

性灵光，随着后来墨家的衰微，几近熄灭。后世的科学家大多注重实用，忽视理性的探索，此实为中国科技史、宇宙学上的损失。

相传还有位学者，列子，在其著作《列子·汤问》中，采用讲故事、二人问答的方式，试图以辩证的思维方式来认识宇宙万物，认识浩瀚宇宙的无限性为“无极无尽”、“无始无终”，认识四海之外的广阔和天地无极，表达了初期的宇宙观念。

摘写原文片段：

殷汤曰：“然则物无先后乎？”夏革曰：“物之终始，初无极已。始或为终，终或为始，恶知其纪？然自物之外，自事之先，朕所不知也。”

殷汤曰：“然则上下八方有极尽乎？”革曰：“不知也。”汤固问。革曰：“无则无极，有则无尽，朕何以知之？然无极之外复无无极，无尽之中复无无尽。无极复无无极，无尽复无无尽。朕以是知其无极无尽也，而不知其有极有尽也。”

译成白话文：

殷汤又问：“这样说，事物的产生就没有先后之分了吗？”夏革回答：“事物的开始和终结，本来就没有固定的准则。开始也许就是终结，终结也许就是开始，又如何知道它们的究竟呢？但是如果说物质存在之外还有什么，事情发生之前又是怎样，我就不知道啦。”

殷汤再问：“那么天地八方有极限和穷尽吗？”夏革回答：“不知道。”殷汤一个劲地问。夏革才回答道：“既然是空无，就没有极限，既然有物，就没有穷尽，那么我凭什么知道呢？因为空无得没有极限之

外‘没有极限’也没有，有物得没有穷尽之中连‘没有穷尽’也没有。没有极限又连‘没有极限’也没有，没有穷尽又连‘没有穷尽’也没有。于是我从这里知道空无是没有极限的，有物是没有穷尽的，而不知道它们是有极限、有穷尽的。”

1.1.4 神话传说

从古以来就有许多神话传说在民间流传，体现着普通百姓美好、善良的愿望和对骤变大自然及现状的无奈；同时，也展露出智者对神话传说的加工、整理和提升，隐含着对周围世界及宇宙的推想。像《盘古开天辟地》、《女娲补天》、《夸父追日》、《后羿射日》、《嫦娥奔月》和《牛郎织女》等，都是家喻户晓，流传至今。

《盘古开天辟地》神话传说的起源可以上溯到远古史前，是最早的关于宇宙起源的神话，同时又与人的始祖结合到一块。

在云南沧源发现的岩画群中，有一幅史前人的画像。

据专家考证，这幅岩画为2万年前原始人的作品，岩画的内容是：1个人头上发出太阳般的光芒，左手握一把石斧，右手拿一个木把，两腿直立，傲视一切。这种形象与盘古立于天地之间，用斧头劈开混沌、开天辟地的传说正相契合。至于人首所呈现的太阳形状，则是反映了原始先民对太阳神的崇拜，也是对盘古把温暖送给人间的希望祈盼。对此，我们完全有理由把它看作是盘古神话的原始因子。据此信息，盘古神话在2万年前就已经诞生。

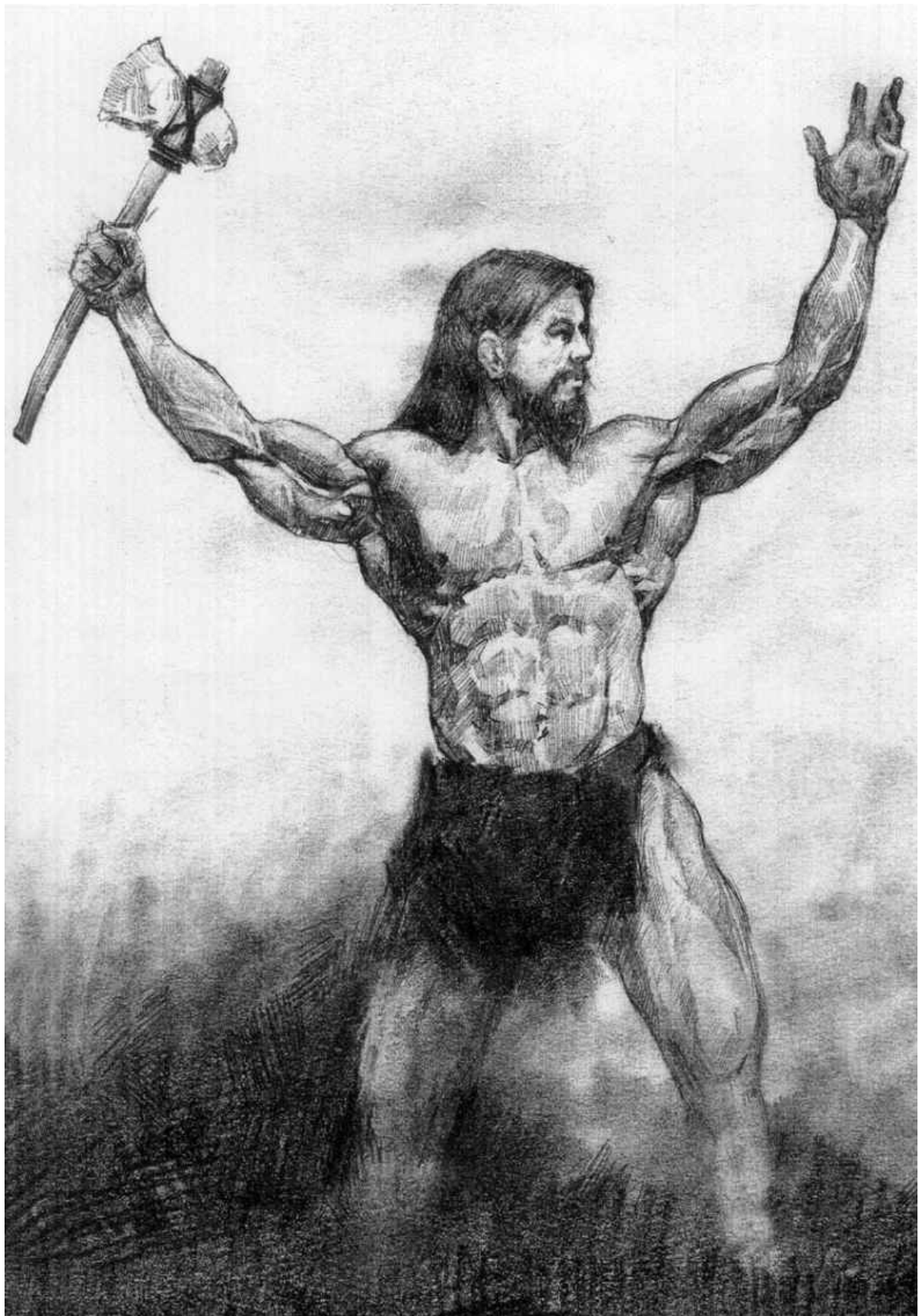
有“活的化石”之称的原始神话，不单纯是人类童年社会先民凭原始思维创作的“口头文学”，也具有原始先民向后世传递史料信息的意义。

而古石初画、初文古字则是原始神话的载体。往往，一个古代石刻、原始符号就能充分地反映出原始先民的文化思维活动，为我们揭示早期人类文化思维活动提供最直接、最可靠的素材。在我国出土的青铜器时代的一个方鼎上，有一个奇特别致的符号。它所表达的意思是什么呢？根据考古学者对甲骨文中“盘”字的正确认识和现存于中国、又在世界上多处出现的十字崇拜中“十字”的原始含义，可以认定，符号的两边是“盘”字的初文简刻，中间的空心十字图案，乃是崇拜的上神，是代表神祖的符号，这个符号应当念为“盘古”。这就像中国文字中的单纯词“尴尬”、“囹圄”等一样是不能分开使用的，只不过这个符号把“盘古”二字合而为一了。这种简朴古拙的做法倒也符合万物起始于简的道理。此符号虽见之于青铜器之上，但不可以说它就产生于铜器时代。其实，它在铜器时代之前就已经产生。由于生产力的限制，人们只能把这一神圣的符号记在心里。而进入铜器时代之后，人们才有能力把它形之于铜器。在史前社会，既然有盘古之名的记载，理应有盘古之事。盘古神话在史前社会的存在，已经是确凿的。

据古籍记载：成都玉堂石室的石刻壁画上刻有自盘古、三皇五帝以来的贤人图像。石刻壁画的创作年代，经考证分析约为公元前8世纪的春秋时期。由此，我们可以了解盘古在春秋时期就已经受到人们的尊崇与敬奉。其形象被刻之于壁，位列三皇五帝之前，足见其影响之深之大，任何人无可比拟。盘古神话在这一时期，于民间广泛流传，已是不争的事实！

传说在天地还没有开辟以前，宇宙就像大鸡蛋一样混沌一团。有个叫做盘古的巨人在这个“大鸡蛋”中一直酣睡了约18 000年后醒来，发现周围一片黑暗。盘古张开巨大的手掌向黑暗劈去，一声巨响，“大鸡蛋”碎了。千万年的混沌黑暗被搅动了，其中又轻又清的东西慢慢上升

并渐渐散开，变成蓝色的天空；而那些厚重混浊的东西慢慢下降，变成了脚下的土地。盘古站在这天地之间，非常高兴。但他很怕天地再合拢起来，还变成以前的样子，于是就用手撑着青天，双脚踏着大地，让自己的身体每天长高一丈。随着他的身体增长，天每天增高一丈，地每天加厚一丈。这样又过了十万八千年，天越来越高，地越来越厚，盘古的身体长得有九万里那么长了。



绘画：张京

盘古凭借着自己的神力，终于把天地开辟出来了。可是盘古也累死了。他临死前，嘴里呼出的气变成了四季飘动的云；声音变成了天空的雷霆；左眼变成了太阳，右眼变成了月亮；头发和胡须变成了夜空的星星；身体变成了东、西、南、北四极和雄伟的三山五岳；血液变成了江河；筋脉变成了道路；肌肉变成了农田；牙齿、骨骼和骨髓变成了地下矿藏；皮肤和汗毛变成了大地上的草木；汗水



盘古

变成了雨露。传说，盘古的精灵魂魄也在他死后变成了人类。所以，都说人类是世上的万物之灵。

1.1.5 小结

中国古代宇宙、天文学刚刚起步，脱离不开稚嫩、谬误之处，但仍取得了大量的成果，影响很深远、特点很突出，令人印象深刻，在世界上占有重要位置。人们思维活跃，积极探索未知的世界；具有朴素的天人合一想法，将人本身及周围的事物与整个世界、宇宙的事物联系、融合在一起来考虑和分析；认为应该与自然和宇宙和谐相处，要顺应天意。这样睿智的宇宙思想令人惊叹。并且还重视时令季节，有较完备的天文历法，并不断完善；注重天象记录，积累留下了世界上最完整、系统的天象记载；制作了灵巧的观测仪器；创作了极富想象力的神话传说，并世代流传。

1.2 古埃及、美索不达米亚和古印度的宇宙及天文学

1.2.1 古埃及

古埃及第三王朝到第六王朝（约公元前27世纪—前22世纪）的文化最为繁荣。古埃及对于天文学的重要贡献，都产生在这一时期。名闻世界的金字塔也是在这一时期建造的。据近代测量，最大的金字塔底座的南北方向非常准确，当时在没有罗盘的条件下，必然是用天文方法测量的。最大的金字塔位于北纬 30° 线南边2千米的地方，塔的北面正中有一入口，从那里走进地下宫殿的通道，和地平线恰成 30° 的倾角，正好对着当时的北极星。

埃及人除了知道北极附近的北极星外，从出土的棺盖上所画的星图可以确定，他们认识的星座还有天鹅、牧夫、仙后、猎户、天蝎、白羊和昴星团等。埃及人认星最大的特征是将赤道附近的星分为36组，每组可能是几颗星，也可能是一颗星。每组占据310天的时间，所以叫做“旬星”。当一组星在黎明前恰好升到地平线上时，就标志着这一旬的到来。现已发现的最早的旬星文物属于第三王朝。

合三旬为一月，合四月为一季，合三季为一年，定年长为360天，这是埃及最早的历法。三个季度的名称分别是：洪水季、冬季和夏季。冬季播种，夏季收获。

在古王国时代，一年中当天狼星清晨出现在东方地平线上的时候，尼罗河就开始泛滥。古埃及人根据对天狼偕日升和尼罗河泛滥的周期进行了长期观测，发现了两者之间的相互联系。大约在公元前18世纪时，

他们把一年由360日增加为365日。这就是现在阳历的来源。但是这与实际周期每年仍约有0.25日之差。如果一年年初第一天黎明前天狼星与太阳同时从东方升起，120年后就要相差一个月，到第1461年又恢复原状，天狼星又与日偕出，埃及人把这个周期叫做天狗周，因为天狼星在埃及叫做“天狗”。直到约公元前13世纪，他们确定年长为365.25天，与地球环绕太阳一圈的时间周期相符合。

据近代研究，埃及除了这种民用的阳历外，还有一种为了宗教祭祀而杀羊告朔的阴阳历。在卡尔斯堡纸草书第九号中有这样一条记载：25埃及年=309月=9125日。

从这条记载就可看出：1年=365日，1朔望月=29.5307日，25年中有9个闰月。

埃及人将昼夜各分为12小时，从日出到日落为昼，从日落到日出为夜，因此一小时的长度是随着季节而不同的。为了表示这种长度不等的时时间，埃及人把漏壶的形状做成截头圆锥体，在不同季节用不同高度的流水量来计算时间。

除了圭表和日晷外，埃及还有夜间用的一种特殊天文仪器，名叫麦开特。它的结构很简单：把一块中间开缝的平板沿南北方向架在一根柱子上，从板缝中可知某星过子午线的时刻，又从星与平板所成的角度知道它的地平高度。现今发现的麦开特，是约公元前1000年的实物，为埃及现存最古老的天文仪器。

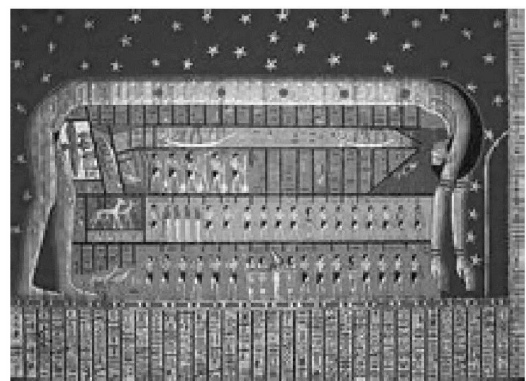
古埃及有许多神话传说，体现出人们最初的宇宙观念，像九柱神的传说：

阿图姆（拉）神是九柱神之首，威力无比的至尊太阳神。他的象征

是一轮金色的圆盘，或是一个中间带有一个点的圆圈的符号。

休是神话中的风神，九柱神之一。他是阿图姆（拉）用自己的精液或分泌物创造出来的。他与自己的妹妹泰芙努特结婚，生下了努特和盖布。他站在自己的儿子——大地之神盖布身上，双手举着女儿努特——天空之神，将他们分开。泰芙努特是雨水之神、生育之神，九柱神之一。

努特是埃及神话中的天神。相较于其他神话中常以男性形象出现的天神，努特是一位女神，她是休与泰芙努特的女儿，九柱神之一。太阳神拉每晚日落后进入她的口中，第二天早晨又从她的阴门中重生。她同时也如此吞咽和再生着星辰。



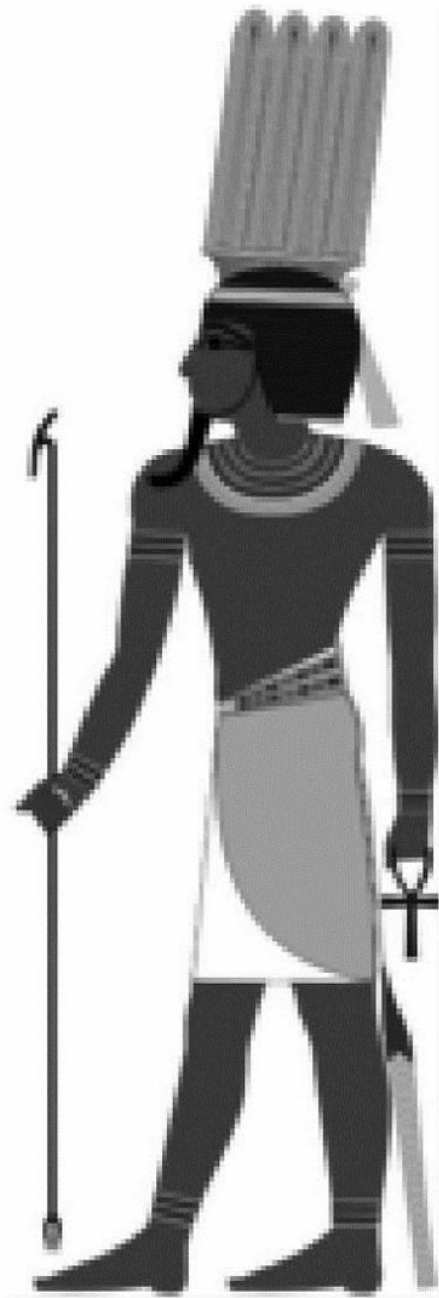
努特

努特同时也是死亡女神，大多数石棺的内壁上都绘有她的形象。法老死后会进入她的身体，不久后便会重生。在艺术作品中，努特的形象是一位被休支撑着，以星辰遮身的裸体女性；在她（天空）的对面是她的丈夫盖布（大地）。

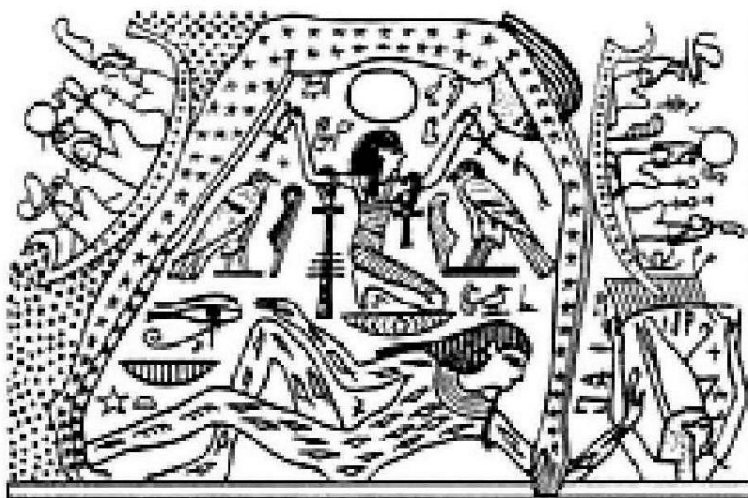
努特与盖布结婚，生下了奥西里斯、伊西斯、赛特和奈芙蒂斯。

盖布（又称作“塞布”或“凯布”）是古埃及的大地之神与生育之神，休与泰芙努特的儿子，九柱神之一。古埃及的这一信仰与世界的其他地区有所不同。在其他的神话中，大地之神往往表现为女神。盖布的形象为鹅头人身，身体呈绿色或黑色。盖布关押着邪恶的人的灵魂，使他们无法进入天堂。

奥西里斯（也作乌西里斯）是埃及神话中的冥王，九柱神之一，是古埃及最重要的神祇之一。他是一位反复重生的神。最终被埋在阿比多斯城，是那里的守护神。



奥西里斯



盖布

伊西斯（希腊语，在埃及语中叫做“阿赛特”）是古埃及的母性与生育之神，九柱神之一。她也是一位反复重生的神。

赛特在埃及神话中最初是力量之神、战神、风暴之神、沙漠之神以及外陆之神。他保护沙漠中的商队，但同时又发起沙暴袭击他们。是九柱神之一。他的形象与亚什神（撒哈拉沙漠之神）紧密结合。

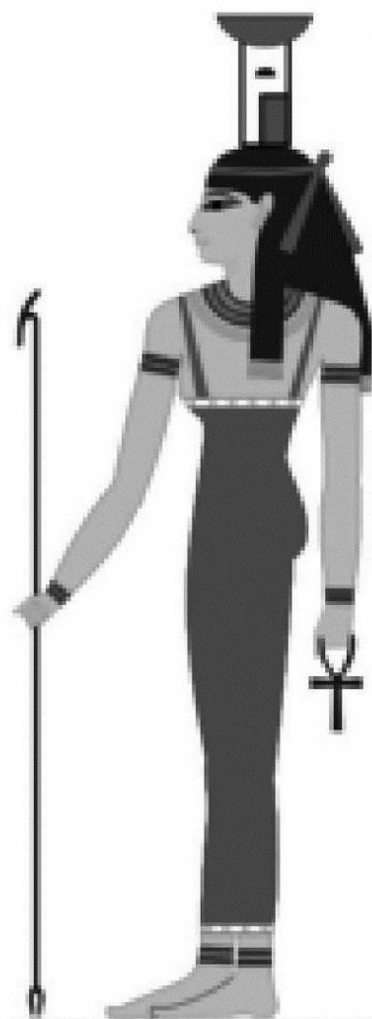
奈芙蒂斯在埃及神话中是死者的守护神，同时也是生育之神。她是九柱神之一。

“奈芙蒂斯”也是对一个家庭中最年长妇女的称呼。在埃及艺术作品中，她的头发看上去与裹尸布相似。她被描绘成头顶一只篮子或一座小房屋，有时是生有双翅的女性，而有时则是风筝、猎鹰、隼或其他鸟类。她常与自己的姐妹伊西斯一同出现在艺术作品中。

1.2.2 美索不达米亚

美索不达米亚是指位于今伊拉克境内的底格里斯河和幼发拉底河流域。是人类古代文明的发源地之一。

这里以数学和天文学的成就为最大。据说在公元前30世纪后期就已经有了历法。当时的月名各地不同。在现在发现的泥板上，有公元前1100年亚述人采用的古巴比伦（约公元前19世纪—前16世纪）历的12个月的月名。因为当时的年是从春分开始，所以古巴比伦历的一月相当于现在的三月到四月。一年有12个月，大小月相间，大月30日，小月29日，一共354天。为了把岁首固定在春分，需要用置闰的办法，补足12个月和回归年之间的差额。公元前6世纪以前，置闰无一定的规律，而是由国王根据情况随时宣布。著名的立法家汉谟拉比曾宣布过一次闰六月。自大流士一世



奈芙蒂斯

（公元前522—前486年在位）后，才有固定的置闰周期，先是8年3闰，后是27年10闰，最后于公元前383年由西丹努斯定为19年7闰制。

古巴比伦的创世神话传说反映出远古时期人们的初始宇宙观念。

关于神话来源：

著名史诗《埃努玛·埃立什》（又称《咏世界创造》）主要汇集了

苏美尔民族的创世思想，着重歌颂地神埃阿之子、主神玛尔都克的事迹。这首诗约1000行，成书于公元前15、14世纪，后经学者从七块泥板中考据整理出来，故又称“七块创世泥板”，它是历史上最早关于创世神话的题材之一。

《埃努玛·埃立什》：

相传太古之初，世界一片混沌，没有天，没有地，只有汪洋一片海。海中有一股咸水，叫做“提亚玛特”；还有一股甜水，叫做“阿普苏”。它们分别代表阴阳两性，在汪洋中不断交汇，生出几个神祇。在生到安沙尔和基沙尔时，他们又生出天神安努和地神埃阿，于是宇宙出现了最初的几代神灵。随着神灵逐渐增多，众神发生争端，提亚玛特和阿普苏日益感到自己的势力在缩小，于是他们决定惩治众神。可是阿普苏并不满意提亚玛特的计划，决心将众神赶尽杀绝。当众神得知这一秘密消息后，便在埃阿神带领下，杀了阿普苏，埃阿神也因此成了众神之首。不久，埃阿神喜得贵子玛尔都克。他生来便与众不同，浓眉大眼、身强力壮，埃阿神又赋予他一切智慧和力量。后来阿普苏的儿子为报父仇，开始向天地神挑战，提亚玛特也前去助阵。天神与之交锋初战告负，决定让玛尔都克一展威风。玛尔都克欣然应允，并做了众神的统治者。他不负众望、英勇作战，一举歼灭来犯者，并亲手切断提亚玛特的腰身，用她的上身筑成苍穹，用她的下半身造出大地。而后他又杀死了提亚玛特的一个辅助神，用他的血造出了人类，并规定人的天职便是侍奉众神。这样玛尔都克终于建立起巴比伦王国，他则成为天国之主、众神之王。

神话的内涵：

这个神话故事是巴比伦文学中较有代表性的作品，它不仅表现了巴

比伦人对创世、人类起源问题的关心、对自然的崇拜，也反映了两河流域国家政治的统一，宗教由多神教向一神教的转变，还表明巴比伦社会从母权制向父权制的过渡，体现了原始社会向奴隶制转变的历史进程。在诗中，提亚玛特代表了阴性世界，她不满众神的强大，欲惩治诸神；代表阳性世界的埃阿神不畏先辈的威力，先斩后奏，夺取王位。埃阿之子玛尔都克继承父业，成为阳性世界的首领。他勇猛顽强、不屈不挠，经过殊死搏斗，终于战胜神母提亚玛特，体现了阳性的刚强和伟大。这个故事与古希腊神话中地母盖亚和众神之主宙斯的故事有些相似，它表现了历史在不断向前迈进的过程，反映了巴比伦王国在两河流域不断统一强大的现实，以及中央集权的政治体制和王权神授的宗教观念。

1.2.3 古印度

印度是世界文明古国之一。印度的天文学起源很早，由于农业生产的需要，印度很早就创立了自己的阴阳历，例如在《梨俱吠陀》中就有关于十三月的记载。

印度上古文献全无年代的记载，要确切地断代是困难的。因此学者们往往借助于天象资料研究历史年代。有些人将吠陀定在公元前25年左右，将梵书定在公元前12世纪，将《吠陀支节录——天文篇》定在梵书之后。但也有人把它们推迟到公元前5世纪前后。

《鹳鹄氏梵书》将一年分为春、热、雨、秋、寒、冬六季；还有一种分法是将一年分为冬、夏、雨三季。《爱达罗氏梵书》记载，一年为360日、12个月，一个月为30日。但实际上，月亮运行一周不足30日，所以有的月份实际不足30日，印度人称之为“消失一个日期”。大约一年要消失5个日期，但习惯上仍称一年360日。印度古代还有其他多种历日

制度，彼此很不一致。在印度历法中还有望终月和朔终月的区别，望终月是从月圆到下一次月圆为一个月；朔终月以日月合朔到下一个合朔为一个月。两种历法并存，但前者更为流行。

印度月份的名称以月圆时所在的星宿来命名。对于年的长度则用观察恒星的偕日出来决定。《吠陀支节录——天文篇》已发明用谐调周期来调整年、月、日的关系。一个周期为5年、1830日、62个朔望月，一个周期内置2个闰月。1个朔望月为29.516日，1年为366日。公元1世纪以前大约一直使用这种粗疏的历法。

为了研究太阳、月亮的运动，印度有27宿的划分方法。它是将黄道分成27等分，称为“纳沙特拉”，意为“月站”。27宿的全部名称最早出现在《鹧鸪氏梵书》。当时以昴宿为第一宿。在史诗《摩诃婆罗多》里则以牛郎星为第一宿，后来又改以白羊座 β 星为第一宿。这个体系一直沿用到后近代。印度27宿的划分方法是等分的，但各宿的起点并不正好有较亮的星，于是他们就选择该宿范围内最高的一颗星作为联络星，每个宿都以联络星的星名命名。印度也有28宿的划分方法，增加的一宿位于人马座和天鹰座间，名为“阿皮季德”，梵文意为“麦粒”宿。

印度天文学在历法计算和宇宙理论上自具特色。令我们惊讶的是，这种科学在古印度是如此先进，古印度天文学家已经记录了星辰的变化，与今天的太空无异。太阳是宇宙（太阳系）的中心，地球的周长是5000瑜伽那斯。1瑜伽那斯等于7.2千米，古代印度估计已经接近真实。但他们不重视对天体的实际观测，因而忽视了天文仪器的使用和制造，在一个很长的时期内仅有平板日晷和圭表等简单仪器。

古印度的神话传说同样反映出人们初始的宇宙观念。创造神梵天在吠陀神话中的身份是祈祷神，而在印度教神话中的正式名字是“梵天”，

并且直接和宇宙最高意志“梵”联系起来。一般认为宇宙中的一切都来自于梵天，而毁灭时又重新归于梵天。梵天就是宇宙最高意志的人格化体现，在印度教神话中被认为是宇宙的创造者，也被称为“世界之主”。但是实际上，梵天受到的崇拜远不及其他两位大神。鉴于世界已经为梵天所创造，人们的注意力自然更多地转移到了毗湿奴和湿婆，亦即维护和毁灭两种力量对于世界权力的争夺上。另外梵天也因为一些所作所为，导致其声誉的降低和信徒的减少——这和《世界·印度篇》的故事主线有直接关系。在整个印度，毗湿奴和湿婆的神庙遍布各地，而专门供奉梵天的神庙却只有一座，那就是普什卡的梵天庙。神话中的梵天为红肤色，四首、四臂，坐骑为天鹅。他在被吸收进入佛教成为护法神后，被称为“大梵天”。

关于梵天的神话传说：

宇宙刚开始的时候，一无所有。首先生产出来的是浩浩荡荡、一望无际的水。水之后，火生成了。在熊熊大火的热力作用下，水中冒出了一个金黄色的蛋。这个蛋在水中漂流了几万年，终于有一天，蛋壳破裂，从中诞生了宇宙万物的始祖——梵天。梵天将蛋壳一分为二，上半部分成了苍天，下半部分成了大地。然后创世之神又在水中开辟了陆地，确定了东南西北的方向，奠定了年月日时的概念。这样宇宙正式形成了。

宇宙形成了，可梵天却发现整个世界除了自己以外，再也没有其他生物，未免感到孤独、寂寞。他心想：我为什么不可以生出后代呢？于是，他马上生出6个儿子，也就是6位伟大的造物主：老大摩里质，生自梵天的心灵；老二阿底利，出自梵天的眼睛；老三安吉罗，出自梵天的嘴巴；老四布罗斯帝耶，出自梵天的右耳；老五布罗诃，出自梵天的左耳；老六克罗图，出自梵天的鼻孔。老大摩里质生了儿子，即著名的仙

人加叶波，他创造出了天神、妖魔、人类、禽兽以及其他生物。老二阿底利也生出了自己的儿子达摩，达摩是正义之神。三儿子安吉罗是安吉罗仙人家族的鼻祖，祭主等大仙就是这个族系中的长者。后来，梵天又用右脚大拇指生出了第七个儿子达刹，左脚趾生出一个女儿毗丽妮，意即夜晚。达刹与毗丽妮后来结为夫妻，一口气生了约50个女儿。其中11个嫁给了加叶波，27个嫁给了月神苏摩，这27个即天上的27个星座，还有10个嫁给了达摩。

达刹的大女儿叫做“蒂提”，是仙人加叶波之妻，也是巨妖底提耶族的母亲。二女儿擅奴，是巨妖擅那婆族的母亲。三女儿阿底提生了12个威武无敌的儿子，他们都是伟大的天神，像海神婆楼那、雷神因陀罗、太阳神苏里耶，而小儿子毗湿奴更是声名赫赫。

蒂提和擅奴的儿子一般称作“阿修罗”，而阿底提的儿子称为“天神”。阿修罗与天神之间经常为争夺对宇宙的控制权而发生战争。他们是势不两立的仇敌。

1.2.4 小结

在远古时代，古埃及、古巴比伦和古印度都曾经创造出辉煌文明，在宇宙及天文领域非常突出。天象观测及记载、完备的历法、想象丰富的创世神话传说和无所拘束的诗篇，都令人赞叹，给后人留下了珍贵的历史遗存。但是，随着岁月的流逝，这三大古文明却失落了，留下许多难解的谜团，至今仍有不少学者在那里探求。巨大的天灾、凶残的瘟疫、外族的入侵，这些会不会是他们失落的原因呢？不过我们总是认为，自身的内部因素估计是最主要的，值得深思。

1.3 古希腊、罗马的宇宙及天文学

1.3.1 创世神话与传说

古代西方，在爱琴海区域和小亚细亚半岛西部的海岸地带，由于生产的发展，使得移民增加、城邦形成，孕育了古希腊时期的文明。在公元前9世纪，希腊诗人荷马的著名诗篇《伊利亚特》和《奥德赛》中就记载着在此之前的丰富的神话与传说。

公元前8世纪希腊诗人希肖特的《神谱》一诗写就了天上诸神的谱系，并保留了远古时代的希腊关于宇宙创生的传说。诗中有一段讲述了天地形成的过程。大意是说，在不知年代的远古，一张张开的大嘴首先出现，这就是混沌初辟。接下来出现的是胸膛宽阔的大地。第三个出现的是爱神爱罗丝。从这张大嘴里生出了阴间和黑夜，它们由爱而结合，黑夜怀胎后产生出明亮的天空和白昼，大地则诞生出有星星的天穹。天穹笼罩着大地，使大地永不动摇。大地又生出高山和大海，山上居住着喜爱山林的半神半人的天神们。

这段有关宇宙创生的神话，包含着神人同形同性这一希腊神话的最大特点。古代人们认为地上生物的滋生繁衍、人在社会中的相互关系等，都可以类推到天上。史前社会里有各民族共同推举的具有权力的领袖，天上就相应会有主宰的神。希腊神话里就有住在奥林匹斯山上的一位主神宙斯和其他神，他们具有无比的威力。地上有战争，希腊神话里同样有关于宙斯和泰坦神族的激烈战事。这些就是宇宙创生神话的由来，也是宇宙演化观念的原始阶段。由于远古社会生产力低下，人对外部自然和自身的认识能力受到很大限制，只能在其所处时代的条件下进

行认识。而且，这些条件达到何种程度，人们便认识到何种程度。

神话在宇宙理论的发展过程中起到了启蒙作用，有积极意义。但当神话被僧侣们利用后就变成了反动的宗教神学，它就只能起着阻碍科学理论发展的有害作用了。后来，在中世纪的欧洲，由于宗教神学势力统治了那时的思想领域，唯物论及辩证法学派宇宙理论和观念的发展受到很大压制。

随着生产力的发展、社会的变革、科学技术的进步、思想的革新，人类顺应自然、认识宇宙的能力日益增强，新的、更高级的神话幻想又会不断产生出来。这是一个永无止境的过程。

1.3.2 关于宇宙的理论

对于宇宙本源的认识，古希腊、罗马的学者进行了卓有成效的探索，形成了早期的理论。

泰勒斯是公元前7—前6世纪的古希腊时期的思想家、科学家、哲学家，是希腊最早的哲学学派——米利都学派（也称“爱奥尼亚学派”）的创始人。他是古希腊七贤之首，西方思想史上第一个有名字留下来的哲学家，是“科学和哲学之祖”。

泰勒斯出生于希腊繁荣的港口城市米利都，曾游历埃及，跟当地祭司学习；曾利用日影来测量金字塔的高度，准确地预测了一次日食；数学上的泰勒斯定理以他命名；他对天文学亦有研究，确认了小熊座，指出其有助于航海事业；同时，他是首个将一年的长度修订为365日的希腊人，他规定一个月为30天；他还第一个测定了太阳从冬至到夏至的运行，并曾估计了太阳及月球的大小。

泰勒斯试图借助经验观察和理性思维来解释世界，是古希腊第一个提出“什么是万物本原”这个哲学问题的人。泰勒斯首创理性主义精神、唯物主义传统和普遍性原则，是理性主义的开端，被称为“哲学史上第一人”。他是个多神论者，认为世间充斥神灵。泰勒斯对希腊哲学产生过重要的影响。

泰勒斯认为“水为万物之源，……一切事物皆营养于润湿，而水为润湿之源。”他第一个提出物质世界的永恒性和统一性；他还是第一个撰写谈论自然的论文的人；据传泰勒斯之前几何学命题都是不加证明而自明的，泰勒斯则证明圆的直径把圆分成的两个面积是相等的，等腰三角形的两个底角是相等的。他是把逻辑推理法引入数学的先驱者。

泰勒斯的继承者们提出宇宙万物的始基是无限，认为“无限变换其部分，而全体则常住不变”。“对立物蕴藏在基质之内，基质是一个无限体，从这个无限体中分离出对立物”，“对立物就是热和冷、湿和干等”。也有的学者提出空气在水之先，是一切物体最单纯的基质。由于基质的稀薄、浓厚不同而形成不同的实体，如火、风、云、水、土和石头等。他们认为“永恒的运动使这些变化发生”；“太阳里有火，星辰具有火的性质，有些也包含具有土的性质的物体，这些物体都为同一运动所牵引着”。

还有的学者提出一个唯一、能动、有限的始基叫做火，火是实体，一切都是火的转换。世界包括一切整体，它不是由任何神或任何人所创造的，它过去、现在和将来都是按规律燃烧着，按规律熄灭着的活火。“一切皆流，物无长住。濯足长流，举足入水，已非前水”。有的提出物质的第一性要素是无限的质的多样性，这些要素以各种不同方式结合成一切存在物。“万物皆如水火，各由相似‘微分’所组成，故生灭只是许多微分之聚散，而各个微分则永恒存在”。组成万物的“微分”有聚有

散，这个概念可称为原子论的先声。他们对天体的性质也提出推测：太阳是一团燃烧的物质；月亮的光借自太阳，月亮上面有山有水，有人居住；彗星是由发出火焰的游星聚集在一起，等等。这些观点各有区别，但都体现出一种原始、自发的唯物论和朴素的辩证法。它在萌芽时期就十分自然地把自然现象的无限多样性的统一看作是不言而喻的，并且在某种具有固定形体的东西中、某种特殊的東西中去找这个统一。

稍后，古希腊出现了原子论学派。大约在公元前5世纪留基伯和他的学生德谟克里特创立了古典原子论。他们认为“充满”和“空虚”是宇宙中最基本的元素，把充实的和坚固的东西称为“原子”。他们认为虚空中被许多物体充满着，当这许多物体进入虚空中并相互混合，就形成了世界。换言之，世界是由原子形成的。他们进一步推测了原子形成世界的过程：“它们聚集在一起，形成一个旋涡，由于旋涡运动，它们彼此冲撞，并按照各个方向转动，这样就相互分开，而相似的物体则结合起来。”留基伯第一个用朴素的唯物主义思想粗略地说出了宇宙形成的动力学过程。德谟克里特也认为“原子在整个宇宙中通过一种旋涡运动而运动着，并因此而形成一些复合物：火、水、气、土”。

原子学派还讨论了原子的大小和形状。留基伯推测原子像微粒，“这些微粒，因为太小，所以是看不见的”。德谟克里特则认为原子在大小和数量上都是无限的，还认为“太阳和月亮是由同样的原子构成。这些原子是光滑和圆的”。虽然他们在原子的细节上有所区别，但是都认为原子是坚硬而不可分的，是物质世界的始基。物质世界的始基概念发展到此时，有了原子的名称，还进一步讨论到它的形状、大小和动力学性质。另外，还有学者认为一切物质的多样性要归结为四种元素，即火、气、水、土。这个关于自然界的四种元素的学说，许多世纪以来一直保存在古希腊、罗马和中世纪哲学中。

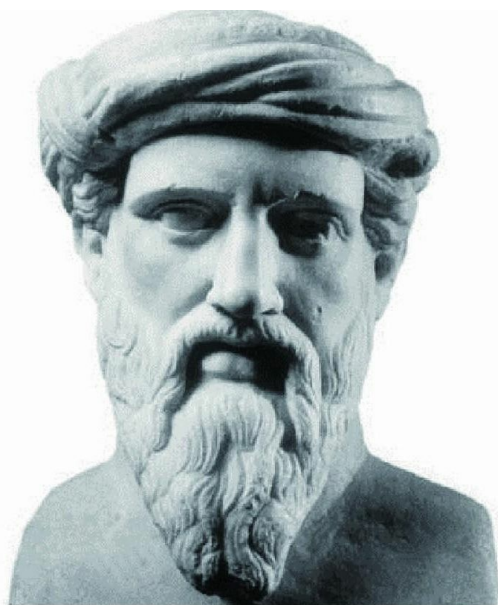
综上所述，这些古希腊学者对宇宙有一个共同的认识，即宇宙以物质为基础——始基，包括“无限”说的始基。宇宙间化生万物的，不是别的，正是在空间上无限、时间上无穷、自身也是无限的东西。在对宇宙本源的认识上，从一开始就以物质来解释世界；并进一步试图解决单一物质和自然界中具体对象的相互关系，使他们朴素的唯物论和辩证法得到发展。这样做，有利于人们分析自然现象、探求自然规律。

对世界的物质性的认识必然会导致否认超自然的东西存在，导致世界不是神所创造，宇宙不是神所治理的无神论断。具有自由意志的神在古希腊唯物论与辩证法学者的宇宙论中，是没有其地位的。尽管也有世界充满着神灵之说，如“神就是永恒的流转的火”等说法。但是，他们所说的神灵或灵魂是和有意志、能造人祸福的神并不相同的。这些所谓的神或灵魂，实质上是在隐晦地指示自然界具有自己的规律性。灵魂是原子的某种集合，这“是天才的猜测，是为科学而不是为僧侣主义指示途径的路标”。

然而，也有一些学者持不同的观点。大约在公元前6世纪，意大利南部出现一个埃利亚学派，他们“假定了一个唯一的始基，把整个存在看成是唯一的東西，认为它既不是无限的，也不是有限的；既不是运动的，也不是静止的。……唯一的宇宙是神”。他们否认感性经验对认识的意义，后来成为唯心论的本源之一。但他们中有人正确地指出化石乃是地球上经历周期性洪水泛滥的明证，这是值得一提的。

还有一个毕达哥拉斯学派。其创始人毕达哥拉斯（约公元前582—前500年）原为爱奥尼亚的学者之一，在埃及等古国游历之后，移居到意大利南部，逐步形成毕达哥拉斯学派，在古希腊、罗马有过巨大影响。他第一个把“数”的哲学概念引入到宇宙本源中去，认为宇宙万物的始基是“一元”，“一元”产生出“二元”。一元是原因，二元是从属于一元

的不定质料，从完满的一元与不定的二元中产生出各种数来。从数产生点，从点产生线，从线产生平面，从平面产生立体，从立体产生出感觉所及的一切物体，产生出四种元素：水、火、土、空气。这四种元素以各种不同的方式互相转化，于是创造出有生命和精神的球形世界。他从宇宙和谐原理出发，认为一切立体图形中最美好的是球形，一切平面图形中最美好的是圆形。而整个宇宙是一个和谐体系，行星运动轨道为圆形，地球



毕达哥拉斯

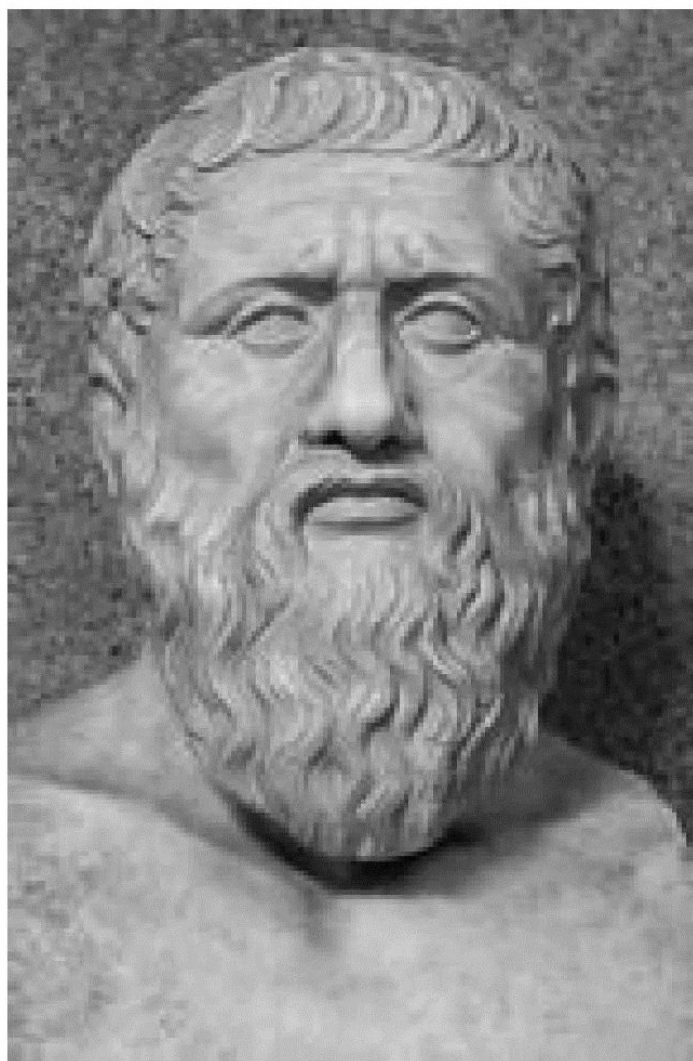
形状是球形的。但是，他把抽象概念的数看作是宇宙万物，包括生命在内的始基，数为万物之源。一切物体都是从数里产生出来的。这就是谬误、违背实际，是唯心的了。毕达哥拉斯学派明白地承认数的基本元素是一切存在的基本元素。因为10是一个完整的数目，所以认为天体的数目也应是10个，不能多也不能少。而当时人们自认为看到9个天体：水星、金星、火星、木星、土星、太阳、月亮、地球和银河，于是他们就编造出第10个天体，即所谓“对地”，更进一步把数推向神秘莫测的境界。他们认为数的某一种特性是正义，另一种特性是灵魂和理性，还有一种特性是机会，一切无不如此。毕达哥拉斯本人还是灵魂轮回说的创始者。他认为太阳、月亮和其他星体都是神灵之物，人类和神灵是亲戚，语言就是灵魂的嘘气，等等。

毕达哥拉斯及其学派把数引入天文学，对天文学的发展做出过贡献，认为宇宙星球及地球是球形的，这是很有见地的，是应该肯定的。但他们认为万物皆数，以为数先于物质而存在，实质上把数绝对化、神

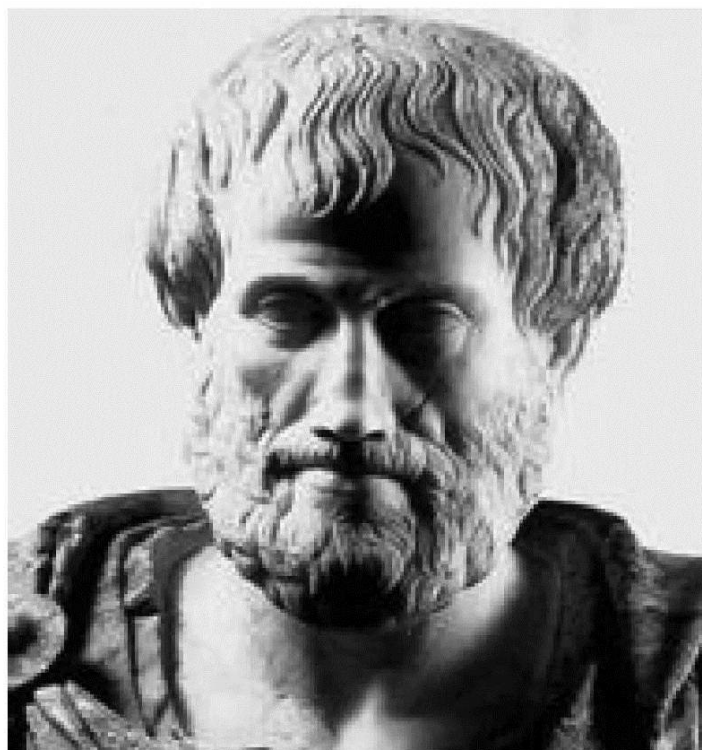
秘化了。在这个基础上，以数来论证天体，就得到天体必须凑足10个数目的谬说，这就走向神秘主义了。数是从感性东西中抽象出来的，这是应当肯定的一面，但“一切抽象在推到极端时都变成荒谬，走向自己的反面”。



苏格拉底



柏拉图



亚里士多德

柏拉图学派的创始人柏拉图（约公元前427—前347年）是著名的古希腊哲学家，他写下了许多哲学的对话录，并且在雅典创办了著名的学院。这所学院成为西方文明中最早的有完整组织的高等学府之一，后世的高等学术机构也因此而得名。柏拉图是苏格拉底的学生，也是亚里士多德的老师，他们三人被广泛认为是西方哲学的奠基者，在宇宙及天文学领域也提出了相应的理论。

柏拉图企图使天文学成为数学的一部分。他认为：“天文学和几何学一样，可以靠提出问题和解决问题来研究，而不去管天上的星界。”柏拉图认为宇宙初始是没有区别的一片混沌。这片混沌的开辟是一个超自然的神，其活动的结果。依照柏拉图的说法，宇宙由混沌变得秩序井然，其最重要的特征就是造物主为世界制定了一个理性方案。关于这个方案付诸实施的机械过程，则是一种想当然的自然事件。

柏拉图的宇宙观基本上是一种数学的宇宙观。他设想宇宙初始有两种直角三角形，一种是正方形的一半，另一种是等边三角形的一半。从这些三角形就合理地产生出四种正多面体，这就组成四种元素的微粒。火微粒是正四面体，气微粒是正八面体，水微粒是正二十面体，土微粒是立方体。第五种正多面体是由正五边形形成的正十二面体，这是组成天上物质的第五种元素，叫做以太。整个宇宙是一个圆球，因为圆球是对称和完善的，球面上的任何一点都是一样。宇宙也是活的、运动的，有一个灵魂充溢全部空间。宇宙的运动是一种环行运动，因为圆周运动是最完善的，不需要手或脚来推动。四大元素中每一种元素在宇宙内的数量是这样的：火对气的比例等于气对水的比例，和水对土的比例。万物都可以用一个数目来定名，这个数目就是表现它们所含元素的比例。在柏拉图的著作《蒂迈欧》和《理想国》中，有不少关于以地球为中心的同心球壳的宇宙结构模型的具体记载。地球在同心球壳的中心保持不动，地球的周围被水包围着，厚度约两倍于地球半径；水之外是空气，厚度约为地球半径的5倍；更外一层是火，厚度为地球半径的10倍，在这层的顶部固定着人们所见天空的万千颗星星。从地球中心到那里的距离总共18倍于地球半径。7个行星则在空气层与恒星圈运行着。从位于中心的地球起，次序为月亮、太阳、水星、金星、火星、木星和土星。柏拉图还设计了一个正多面体的宇宙结构模型，称为柏拉图图形。他试图以不同的正多面体把内外两层的同心球壳联系起来。每两个相邻的球壳之间包含一个正多面体，它的角和外球壳的内壁相接触，这个正多面体的面则和内球壳外壁相切。不同面数的正多面体把这些同心球壳联系在一起。柏拉图把正多面体的面数和同心球壳的半径作了神秘主义的解释。即便有些神秘色彩，但这个宇宙结构模型毕竟启发了研究宇宙结构的后继者。

柏拉图是西方客观唯心主义的创始人，认为任何一种哲学要能具有

普遍性，必须包括一个关于自然和宇宙的学说在内。柏拉图试图掌握有关个人和大自然永恒不变的真理，因此发展一种适合并隶属于他的政治见解和神学见解的自然哲学。

柏拉图认为，自然界中有形的东西是流动的，但是构成这些有形物质的“形式”或“理念”却是永恒不变的。柏拉图指出，当我们说到“马”时，我们没有指任何一匹马，而是称任何一种马。而“马”的含义本身独立于各种马（“有形的”），它不存在于空间和时间中，因此是永恒的。但是某一匹特定、有形、存在于感官世界的马，却是“流动”的，会死亡和腐烂。这可以作为柏拉图的“理念论”的一个初步解说。

柏拉图认为，我们对那些变换、流动的事物不可能有真正的认识，我们对它们只能有意见或看法。我们唯一能够真正了解的，只有那些能够运用理智来了解的“形式”或者“理念”。因此柏拉图认为，知识是固定的、肯定的，不可能有错误的知识。但是意见是有可能错误的。

柏拉图在其著作中对许多领域都提出了自己的主张，特别著名的有《理想国》、“柏拉图式爱情观”等。

在柏拉图的《理想国》中，有一个著名的洞穴比喻来解释理念论：有一群囚犯在一个洞穴中，他们手脚都被捆绑，身体也无法转身，只能背对着洞口。他们面前有一堵白墙，身后则燃烧着一堆火。在那面白墙上，他们看到了自己以及身后到火堆之间的事物的影子，由于他们看不到任何其他东西，所以以为影子就是真实的东西。最后，一个人挣脱了枷锁，并且摸索出了洞口。他第一次看到了真实的事物，于是他返回洞穴并试图向其他人解释，那些影子其实只是虚幻的事物，并为他们指明了光明的道路。但是对于那些囚犯来说，那个人似乎比他逃出去之前更加愚蠢，并向他宣称，除了墙上的影子之外，世界上没有其他东西了。

柏拉图利用这个故事来告诉我们，“形式”其实就是那些阳光照耀下的实物，而我们的感官世界所能感受到的不过是那白墙上的影子而已。我们的大自然比起鲜明的理性世界来说，是黑暗而单调的。不懂哲学的人能看到的只是那些影子，而哲学家则在真理的阳光下看到外部事物。但是另一方面，柏拉图把太阳比作正义和真理，强调我们所看见的阳光只是太阳的“形式”，而不是实质；正如真正的哲学道理、正义一样，是只可见其外在表现，而其实质是不可言说的。

柏拉图的《理想国》还向我们描绘出了一幅理想的乌托邦的画面，设计了一幅正义之邦的图景：国家规模适中，以站在城中高处能将全国尽收眼底，国人可以彼此面识。柏拉图认为国家起源于劳动分工，因而他将理想国中的公民分为治国者、武士、劳动者三个等级，分别代表智慧、勇敢和欲望三种品性。治国者依靠自己的哲学智慧和道德力量统治国家；武士们辅助治国，用忠诚和勇敢保卫国家的安全；劳动者则为全国提供物质生活资料。三个等级各司其职，各安其位。在这样的国家中，治国者均是德高望重的哲学家，只有哲学家才能认识理念，具有完美的德行和高超的智慧，明了正义之所在，按理性的指引去公正地治理国家。治国者是少部分管理国家的精英。他们可以被继承，但是其他阶级的优秀儿童也可以被培养成治国者，而治国者中的后代也有可能被降到普通人民的阶级。治国者的任务是监督法典的制定和执行情况。为达到该目的柏拉图有一整套完整的理论。他的理想国要求每一个人在社会上都有其特殊功能，以满足社会的整体需要。但是在这个国家中，女人和男人有着同样的权利，存在着完全的性平等。政府可以在为了公众利益时撒谎。每一个人应该去做自己分内的事而不应该打扰到别人。在《理想国》中，治国者和武士没有私产和家庭，因为私产和家庭是一切私心邪念的根源。劳动者也绝不允许拥有奢华的物品。理想国还很重视教育，因为国民素质与品德的优劣决定国家的好坏。在今天看来，柏拉

图描绘的理想国是一个可怕的极权主义国家。但是“理想国其实是用正确的方式管理国家的科学家的观点”。

柏拉图才思敏捷、研究广泛、著述颇丰。以他的名义流传下来的著作有40多篇，另有13封书信。柏拉图的主要哲学思想都是通过对话的形式记载下来的。在柏拉图的对话中，有很多是以苏格拉底之名进行的谈话，因此人们很难区分哪些是苏格拉底的思想，哪些是柏拉图的思想。经过后世一代代学者艰苦细致的考证，其中有24篇和4封书信被确定为真品。它以理念论为中心，包括宇宙论方面的宇宙生成说、认识论方面的回忆说、伦理观与社会政治观方面的四主德与理想国的学说、美学方面的“摹本”说、探求理念体系的概念辩证法以及教育学说等。它是欧洲哲学史上第一个庞大的客观唯心主义体系，对后世西方哲学的影响极大。

苏格拉底的审判和死刑对柏拉图造成极大的震撼，更让他感到失望和恶心。苏格拉底的审判是一系列对话录中最为着重、也最为一致的事件。柏拉图在许多对话录都曾明确或间接地提起这场审判，或提起这场审判的情节和角色。在《泰阿泰德篇》和《伊壁鸠鲁篇》中苏格拉底告诉大家他必须面临一场不公平的审判。而在《美诺篇》里，阿尼图斯则警告苏格拉底应该避免批评当时的重要人物，以免使自身惹上麻烦，阿尼图斯在《申辩篇》里也是那些联合起诉苏格拉底的人之一。《申辩篇》是苏格拉底的辩护演说，《克力同篇》和《斐多篇》则是在审判定罪后于监狱内的对话。

亚里士多德从18岁到38岁——在雅典跟柏拉图学习哲学的20年，这一时期的学习和生活对他一生产生了决定性的影响。苏格拉底是柏拉图的老师，亚里士多德又受教于柏拉图。在雅典的柏拉图学院中，亚里士多德表现得很出色。

亚里士多德在哲学上最大的贡献在于创立了形式逻辑这一重要的分支学科。逻辑思维是亚里士多德在众多领域建树卓越的支柱，这种思维方式自始至终贯穿于他的研究、统计和思考之中。他在研究方法上，习惯于对过去和同时代的理论持批判态度，提出并探讨理论上的盲点，使用演绎法推理，用三段论的形式论证。他留下了“吾爱吾师，吾更爱真理”的名言。

在宇宙及天文学领域，亚里士多德认为运行的天体是物质的实体；地球是球形的，是宇宙的中心；地球和天体由不同的物质组成，地球上的物质是由水、气、火、土4种元素组成，天体由第5种元素“以太”构成。

亚里士多德是现实主义的鼻祖。不同于他的老师柏拉图以自己假定的理想国衡量现实，他主张从现实的国家出发，防止国家堕落和促进国家发展。他对人性和理性持怀疑态度，主张法治，而法律的来源也不是人的理性或者学者的思考，而是来自于历史 and 传统中为人们所遵循和认知的东西，也就是历史的理性。他对变法和改革持一种十分谨慎的态度。亚里士多德显示了希腊科学的一个转折点。在他以前，科学家和哲学家都力求提出一个完整的世界体系来解释自然现象，他是最后一个提出完整世界体系的人。在他以后，许多科学家放弃提出完整体系的企图，转入研究具体问题。



古希腊雅典学派

1.3.3 天象观测和天文历法

关于可以辨别的恒星和星座的记载，早在现存最早的古希腊文学作品——荷马和赫西俄德的作品中就出现了。在《伊利亚特》和《奥德赛》中，荷马提到了这些天体：牧夫座、毕星团、猎户座、昴星团、天狼星、大犬座。

活跃于公元前7世纪的赫西俄德则在《工作和时日》中提及了大角星。它们传达了一种原始的宇宙学——平坦的大地被一条大洋河所包围。一些恒星会升起和落下（从古希腊人的观点来看，落下即是消失在海洋中）；而其他恒星则是不落的。根据一年中时候的不同，有些恒星会在日出或日落的时候升起或落下。

水星、金星、火星、木星、土星这5颗行星可以裸眼观察到；有时候太阳和月亮也被归类成裸眼行星。由于行星在接近太阳时时常会被太

阳的光芒掩盖，要识别出全部5颗行星需要进行仔细的观察。金星就是这样一个例子。早期的古希腊人认为在傍晚和清晨出现的金星分别是两个不同的天体，直到毕达哥拉斯发现它们其实是同一个行星。那时已有彗星记录，但很不完整。

许多古代历法都以太阳或月亮的运行周期为基础。古希腊历法中也包含这两个周期。然而，同时基于太阳和月亮的周期的阴阳历并不容易编制。一些古希腊天文学家创造出了基于食的周期的历法。

公元前5世纪古希腊学者已测定年的长度为365.2632天，朔望月29.531 92天；创立19年7闰的历法；采用阴阳历。后来，又有学者更精确地确定年长为365.2467天，朔望月29.530 85天。

1.3.4 小结

古希腊罗马是一个引人注目的时代，在宇宙及天文学领域与哲学、数学、物理学、文学、社会学、建筑学等领域一样，同样也取得了巨大的成就，为后世的持续发展奠定了良好的基础，让人赞叹。这丰富、宝贵的知识遗产，至今仍吸引着学者们去研究。我们简略地归纳、总结一下其原因和特点。

古希腊城邦的形成和大量的移民使各种文化融会。学者们注意吸取、学习其他文明古国的先进知识，甚至去那里实地游历，经过理性分析，形成新的自己的观念。

综合多方面的知识，具有各种文化理性因子：毕达哥拉斯创新的数学理性、亚里士多德的逻辑理性和阿基米德的实验理性。特别注重利用高度的几何学成就和严密的逻辑体系来推算宇宙天体的运动及宇宙结构

模型。

思想活跃、开放，氛围宽松、民主，努力探求未知的事物。在探讨天文现象的内在联系和宇宙本源及结构理论方面产生了众多的学派。

强调从大自然本身来解释宇宙中观察到的一切事物。具有朴素、自发的唯物思维和辩证思维。

参考文献

- [1] 江晓原, 钮卫星. 中国天文学史 [M]. 上海: 上海人民出版社, 2005.
- [2] 陈美东. 中国古代天文学思想 [M]. 北京: 中国科学技术出版社, 2007.
- [3] 杨国安. 品悟老子 [M]. 北京: 中国长安出版社, 2012.
- [4] 廖名春. 中国学术史新证 [M]. 成都: 四川大学出版社, 2005.
- [5] 方勇. 墨子 [M]. 北京: 中华书局, 2011.
- [6] 袁珂. 中国古代神话 [M]. 北京: 华夏出版社, 2013.
- [7] 袁珂. 中国神话传说 [M]. 北京: 世界图书出版公司, 2012.
- [8] [法] G. 伏古勒尔. 天文学简史 [M]. 罗玉君, 译. 北京: 中国人民大学出版社, 2010.
- [9] 钟怡阳. 流传千年的埃及神话故事 [M]. 南京: 南京大学出版社, 2013.
- [10] 李明滨. 世界文学简史 [M]. 北京: 北京大学出版社, 2007.
- [11] 百度网. 古印度天文学 [OL]. 2015-03-14.
<http://baike.baidu.com/view/26066.htm>
- [12] 朱维之. 外国文学简史 [M]. 北京: 中国人民大学出版社,

2004.

- [13] [希] 赫西俄德. 工作与时日. 神谱 [M]. 蒋平, 译. 北京: 商务印书馆, 1991.
- [14] [希] 亚里士多德. 形而上学 [M]. 苗力田, 译. 北京: 北京出版社, 2008.
- [15] 北京大学哲学系 / 外国哲学史教研室. 西方哲学原著选读 [M]. 北京: 商务印书馆, 2014.
- [16] 恩格斯. 自然辩证法 [M]. 中共中央马恩列斯著作编译局, 译. 北京: 人民出版社, 1971.
- [17] [希] 柏拉图. 理想国 [M]. 吴天岳, 译. 北京: 北京理工大学出版社, 2010.
- [18] 本书编写组. 最博学的人——亚里士多德 [M]. 北京: 中国国际广播出版社, 2014.

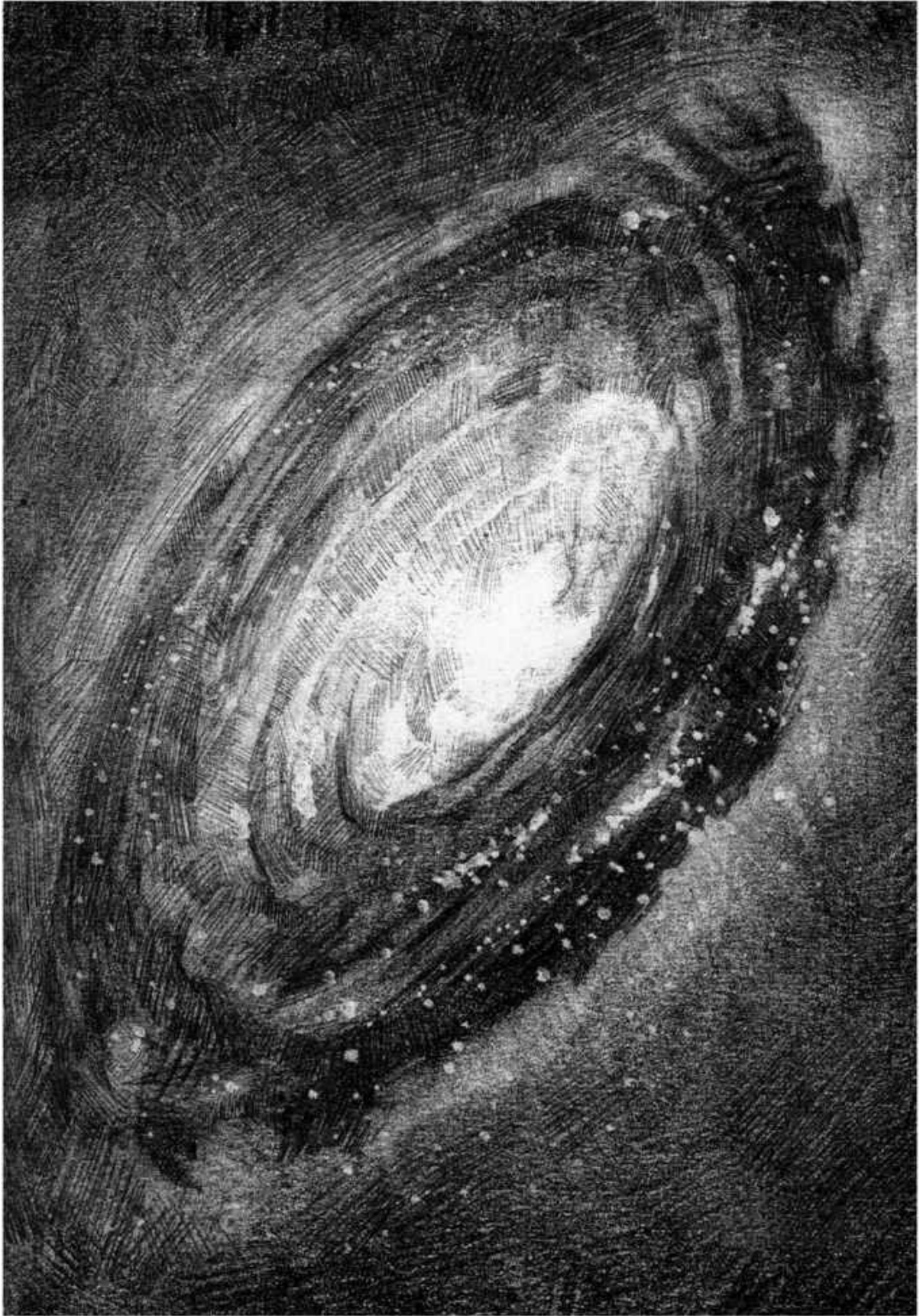
2 宇宙及天文学的继续发展

随着时代的进步，中国古代的天文学不断发展，形成一个较为完备的体系，在历法编制推算、天文仪器制造、大地测量和宇宙观念等方面都有新成就，达到较高水平。西方各学派对宇宙本源和结构展开了争论，出现了最早的日心说推测，但地心说却更为盛行，占有正宗地位。之后进入宗教神学的黑暗时期。

2.1 中国的宇宙及天文学的进展

2.1.1 宇宙观念

宇宙理论浑天说形成于大约公元前3世纪战国时期，于东汉张衡的理论中渐趋完善。在《浑天仪图注》中，张衡谈道：“浑天如鸡子。天体圆如弹丸，地如鸡子中黄，孤居于天内，天大而地小。天表里有水，天之包地，犹壳之裹黄。天地各乘气而立，载水而浮.....天转如车毂之运也，周旋无端。其形浑浑，故曰浑天。”张衡认为，天是一种圆球体，地球如在球中，如同蛋黄在蛋内一般。



绘画：张京

浑天说认为全天恒星都布于一个“天球”之上，而日月星辰则在于“天球”上运行。浑天说不仅是一套宇宙理论，也是一套观测和测量天体视运动的计算体系。以浑天仪为代表的观测仪器为中国古代天文学观测做出了重大贡献。浑天说于张衡理论中成形，历经魏晋时期的论证、改造，再到唐朝对本初子午线的测量，使得浑天说成为西方近代天文学传入之前的正统学说。

同浑天说相类似，宣夜说也是古人提出的一种宇宙学说。《晋书·天文志》说：“宣夜之书亡，惟汉秘书郎郗萌记先师相传云，天了无质，仰而瞻之，高远无极，眼瞀精绝，故苍苍然也。譬之旁望远道之黄山而皆青，俯察千仞之深谷而幽黑。夫青非真色，而黑非有体也。日月众星，自然浮生虚空之中，其行其止皆须气焉。是以七曜（七曜指日、月及金、木、水、火、土五星）或逝或住，或顺或逆，伏见无常，进退不同，由乎无所根系，故各异也。故辰极常居其所，而北斗不与众星同没也；摄提、填星皆东行，日行一度；月行十三度。迟疾任情，其无所系著可知矣，若缀附天体，不得尔也。”

这是关于宣夜说的一段最完整的史料，它包含了有关宣夜说的许多内容。首先，宣夜说起源很早，汉代郗萌（公元1世纪）只是记下了先师传授的东西。第二，宣夜说认为天是没有形体的无限空间，因无限高远才显出苍色。第三，以远方的黄色山脉看上去呈青色，千仞之深谷看上去呈黑色，而实际上山并非青色，深谷并非有实体，以此证明苍天既无形体，也非苍色。第四，日月众星自然浮生虚空之中，依赖气的作用而运动或静止。第五，各天体运动状态不同，速度各异，是因为它们不是附缀在有形质的天上，而是飘浮在空中。

无可否认，这些看法是相当先进的，它同盖天、浑天说本质的不同在于：它承认天是没有形质的，天体各有自己的运动规律，宇宙是无限的空间。这三点即使在今天也是有意义的。或许正因为它的先进思想距离当时人们的认识水平太远，所以不可能为多数人所接受。试想，一个无限的宇宙空间已是难以想象，更何况众多的天体都毫无依赖地飘浮在空中、各自运动呢？在近代科学诞生以后，人们依据万有引力定律和天体力学规律说明了天体的运动，证明了宣夜说的基本观点是正确的，与现代科学理论颇为接近，具有超前的认识水平与思辨性。然而，在古代缺少先进的科学手段加以证明，使得宣夜说只能停留于思想层面，成为一种思辨的假说。且其超越性使得时人无法接受，险些失传。

随着时间的流逝，人们对宣夜说的观点也渐渐淡忘了。唐代学者所著的《晋书·天文志》中保留了宣夜说的唯一资料，才使这一思想得以保存下来。

中国古代的宇宙学理论起源较早，发展却相对迟缓，缺少数学性的理论完善与实验性检验，使得这些理论大多停留在思想理论层面，最终为近代西方天文学理论所取代。

2.1.2 宇宙思想认识

进入春秋战国时期后，我国学术界的思想异常活跃，诸子百家纷纷提出了各自新的学说。其中有代表性的学说包括：《老子》提出的宇宙物质起源学（有生于无），《墨子》进行的光学（小孔成像原理）、机械学研究，《庄子·天下》记载的各派学说（涉及物质的微观结构和宏观结构等多方面的数理逻辑思辨课题，诸如原子论、相对论、运动论、多维空间论等，可惜各学派学者的原著已经失传），《计倪子》记载的

气候经济学（根据木星12年绕太阳一周所引起的降雨量周期变化，提前准备开展相应的经济贸易活动），以及《尸子》记述的平面几何、测量学，《列子·汤问》等篇关于天地结构、宇宙万物、远方异国的种种思考和忧虑（杞人忧天）等。这里，不能不提到屈原和他的思想代表作《天问》。

屈原（约公元前340—前278年），姓屈，名平，字原；汉族，战国末期楚国丹阳人；楚武王熊通之子屈瑕的后代，贵族、士大夫、学者。屈原是中国最伟大的浪漫主义诗人之一，也是我国已知最早的著名诗人、世界文化名人。他创立了“楚辞”这种文体，也开创了“香草美人”的传统，代表作品有《离骚》、《九歌》等。



屈原

屈原早年受楚怀王信任，任左徒、三闾大夫，常与怀王商议国事、参与法律的制定，他观天定法，主张章明法度，楚国国力因此有所增强。但由于自身性格耿直，加之他人谗言与排挤，屈原逐渐被楚怀王疏远。公元前305年，屈原反对楚怀王与秦国订立黄棘之盟，但是楚国还是彻底投入了秦的怀抱。屈原亦被楚怀王逐出郢都，开始了流放生涯。

在流放期间，屈原为后世留下了许多不朽名篇。其作品文字华丽、想象奇特、比喻新奇、内涵深刻，成为中国文学的起源之一。长诗《天问》即体现出屈原深邃的思想内涵，特别是宇宙观念。这里对其第一部分作一介绍。

《天问》（第一部分）原文：

曰：遂古之初，谁传道之？

上下未形，何由考之？

冥昭瞢暗，谁能极之？

冯翼惟像，何以识之？

明明暗暗，惟时何为？

阴阳三合，何本何化？

圜则九重，孰营度之？

惟兹何功，孰初作之？

斡维焉系，天极焉加？

八柱何当，东南何亏？

九天之际，安放安属？

隅隈多有，谁知其数？

天何所沓？十二焉分？

日月安属？列星安陈？

出自汤谷，次于蒙汜。

自明及晦，所行几里？

夜光何德，死则又育？

厥利维何，而顾兔在腹？

女歧无合，夫焉取九子？

伯强何处？惠气安在？

何阖而晦？何开而明？

角宿未旦，曜灵安藏？

.....

译文：

请问远古开始之时，谁将此态流传导引？
天地尚未成形之前，又从哪里得以产生？
明暗不分混沌一片，谁能探究根本原因？
迷迷蒙蒙这种现象，怎么识别将它认清？
白天光明夜晚黑暗，究竟它是为何而然？
阴阳参合而生宇宙，哪是本体哪是演变？
天的体制传为九重，有谁曾去环绕量度？
这是多么大的工程，是谁开始把它建筑？
天体轴绳系在哪里？天极不动设在哪里？
八柱撑天对着何方？东南为何缺损不齐？
平面上的九天边际，抵达何处联属何方？
边边相交隅角很多，又有谁能知其数量？
天在哪里与地交会？黄道怎样十二等分？
日月天体如何连属？众星在天如何置陈？
太阳是从暘谷出来，止宿则在蒙汜之地。
打从天亮直到天黑，所走之路究竟几里？
月亮有着什么德行，竟能死了又再重生？
月中黑点那是何物，是否兔子腹中藏身？
神女女歧没有配偶，为何能够产下九子？
伯强之神居于何处？天地瑞气又在哪儿？
天门关闭为何天黑？天门开启为何天亮？
东方角宿还没放光，太阳又在哪儿匿藏？
.....

这篇包含着作者深层思想结晶的《天问》是其宇宙思想学说的集萃，所问的都是上古传说中不甚可解的怪事、大事，“天地万象之理，存亡兴废之端，贤凶善恶之报，神奇鬼怪之说”，他似乎是要求得一个

解答，找出一个因果。而这些问题也都是春秋、战国以来许多学者所探究的问题，在诸子百家的文章里，几乎都已讨论到。屈子的《天问》则以惆怅迷离的词句，用疑问的语气说出来，这就是屈子所以为诗人而不是“诸子”的缘故。而“天”字的意思，战国时代含义已颇广泛。大体来说，凡一切远于人、高于人、古于人，人所不能了解、不能施为的事与物，都可用“天”来统摄之。对物质界说，又有本始、本质、本原的意思。屈原为楚之宗室重臣，有丰富的学识和经历，以非凡才智作此奇文，颇有整齐百家、是正杂说之意，《天问》的光辉和价值也就很清楚地呈现于读者面前了！

从全诗的结构及内容来看，全诗372句、1553字，是一首以4字句为基本格式的长诗，对宇宙及天文、地理、历史、哲学等许多方面提出了170多个问题。这些问题有许多是在他那个时代尚未解决而他有怀疑的，也有明知故问的对许多历史问题的提问，往往表现出作者的思想感情、政治见解和对历史的总结、褒贬；尤其是对自然所提的问题。从篇首至“曜灵安藏”，这部分屈子问的是天，宇宙生成是万事万物的先决，这便成了屈原问难之始，其中从“遂古之初”至“何以识之”问的是天体的情况，“明明暗暗”四句讲宇宙阴阳变化的现象。第二小节自“圜则九重”到“曜灵安藏”，这一部分是对自然结构提出问题，先对宇宙起源、天体结构和日月星辰运行发问，接下来对大地结构和鲧禹治水、羿射十日等事件发问。事实上，我国古人很早就产生了天有九层的观念，但是最早见诸文字的是《天问》中“圆则九重，孰营度之”、“九天之际，安放安属”。这里涉及宇宙的空间深度和天上物体彼此之间的距离问题。从视觉直观的角度来说，古人所说的九重天大体可以分为如下九个层次：距离大地最近的是云、雾、雷、电（我国古代经常把气象归入天文现象），然后是月亮，接下来是内行星（水星、金星，它们能够出现在太阳之前，形成凌日现象，表明它们比太阳近）、太阳、外行星（火

星、木星、土星）、彗星、亮的恒星、暗的恒星，最遥远的是模糊的星云。

需要说明的是，我国古代天文学非常发达，七八千年前就能够根据四颗恒星判断四季（尧典四星）；4000年前就已经测定出木星12年绕太阳运行一周；最早的日食记录、太阳黑子记录、哈雷彗星记录，以及最早的星表（战国时代编制）都出现在中国。

《天问》表现的是屈原对宇宙本源和构成的探索精神，对传统说法的质疑，从而也看出屈原比同时代人进步的宇宙观、认识论及高尚的人格。《天问》以新奇的艺术手法表现精深的内容，使之成为世界文库中绝无仅有的奇作。

战国、秦汉时期还有一些学者在其著作中（如《墨经》、《尸子》等）对宇宙进行了科学的诠释，“四方上下曰宇，往古来今曰宙”。张衡更是对这种宇宙观念作了精辟的发挥，“宇之表无极，宙之端无穷”。就是说：宇，表示的是空间，其范围是无边无际的；宙，表示的是时间，其延伸是无穷无尽的。

荀子（约公元前313—前238年）在其著作中表达了“天道自然”、“天行有常”、“天人相分”和“制天命而用之”的宇宙思想。

荀子将“天”、“天命”、“天道”自然化、客观化与规律化，见于他的《天论》一文。“列星随旋，日月递炤，四时代御，阴阳大化，风雨博施，万物各得其和以生，各得其养以成，不见其事而见其功，夫是之谓神；皆知其所以成，莫知其无形，夫是之谓天。”

在他看来，天为自然，没有理性、意志、善恶和好恶之心。天是自然天，而不是人格神。他把阴阳、风雨等潜移默化的机能叫做“神”，把

由此机能所组成的自然界叫做“天”。宇宙的生成不是神造，而是万物自身运动的结果。

荀子以为，天不是神秘莫测、变幻不定的，而是有自己不变的规律。这一规律不是神秘的天道，而是自然的必然性，它不依赖于人间的好恶而发生变化。人不可违背这一规律，而只能严格地遵守它。



“天行有常，不为尧存，不为桀亡。应之以治则吉，应之以乱则凶。”天道不会因为人的情感或者意志而有所改变，对人的善恶分辨完全漠然置之。荀子对传统的宗教迷信持批判的态度，认为自然的变化与社会的治乱吉凶没有必然的联系。他认为祭祀、哀悼死者的各种宗教仪式，仅仅是表示“志意思慕之情”，是尽“人道”而非“鬼事”（《礼论》）。

荀子认为自然界和人类各有自己的规律和职能。天道不能干预人道，天归天，人归人，故言天人相分不言合。治乱吉凶，在人而不在天。并且天人各有不同的职能，“天能生物，不能辨物，地能载人，不能治人”（《礼论》）；“天有其时，地有其才，人有其治”（《天论》）。

在荀子看来，与其迷信天的权威，去思慕它、歌颂它，等待“天”的恩赐，不如利用自然规律以为人服务。荀况强调“敬其在己者”，而不要“慕其在天者”。他甚至以对天的态度作为君子、小人之分的标准，强

调人在自然面前的主观能动性，主张“治天命”、“裁万物”、“骋能而化之”的思想。荀子明确地宣称，认识天道就是为了能够支配天道而宰制自然世界。

王充（公元27年—约97年）在《论衡》一书中同样从宇宙观上否定了“天人感应”的“天”，还世界的物质性面貌。不过，《论衡》一书中所描述的宇宙观，是一种自然主义的宇宙观：“天地合气，物偶自生也”、“及其成与不熟，偶自然也”（《论衡·物势》篇）。所以，这种宇宙观只能是人能利用自然，辅助“自然之化”，但终究得听命于自然力的支配。

1. 天自然无为

王充认为天和地都是无意志的自然物质实体，宇宙万物的运动变化和事物的生成是自然无为的结果。他认为万物是由于物质性的“气”自然运动而生成的，“天地合气，万物自生”。生物间的相胜是因为各种生物筋力的强弱、气势的优劣和动作的巧拙不同，并非天的有意安排，天不是什么有意志、能祸福的人格神。



王充

2. 天不能故生人

王充认为天是自然，而人也是自然的产物，“人，物也；物，亦物也”，这样就割断了天人之间的联系。他发扬了荀子“明于天人之分”的唯物主义思想。他说：“人不能以行感天，天亦不能随行而应人”。他认

为社会的政治、道德与自然界的灾异无关，所谓“天人感应”的说法只是人们以自己的想法去比拟天的结果。

3. 神灭无鬼

王充认为人有生即有死。人所以能生，由于他有精气血脉，而“人死血脉竭，竭而精气灭，灭而形体朽，朽而成灰土，何用为鬼？”他认为人死犹如火灭，火灭为何还能有光？他对于人的精神现象给予了唯物的解释，从而否定鬼的存在，破除了“善恶报应”的迷信。

4. 今胜于古

王充反对“奉天法古”的思想，认为今人和古人相“齐”。即认为今人与古人气禀相同，古今不异，没有根据说古人总是胜于今人，没有理由颂古非今。他认为汉比过去进步，汉在“百代之上”，因为汉在“百代”之后。这种见解与“天不变道亦不变”的思想是完全对立的。

2.1.3 飞天的神话传说、牛郎织女和嫦娥奔月

自古以来，我国就流传着许多关于飞天的神话传说。据考古发现，早在战国墓葬的石室壁上，就刻有飞天的图像。敦煌壁画中那动人的飞天形象更是人人皆知。这体现着人们久远的飞天梦想和在太空中自由飞翔的美好愿望。下面讲述两个在民间广为流传的牛郎织女和嫦娥奔月的故事。

“织女”、“牵牛”二词见诸文字，最早出现于《诗经》中的《大东》篇。诗中的织女、牵牛只是天上两个星座的名称，它们之间并没有什么关系，但会引起人们的联想。到了东汉时期，无名氏创作的《古诗十九首》中，有一首《迢迢牵牛星》，从中可以看出，牵牛、织女已是一对

相互倾慕的恋人，不过诗中还没有认定他们是夫妻。在文字记载中，称牛郎、织女为夫妇的，最早应是南北朝时期编纂的《文选》，其中有一篇《洛神赋》的注释中说：“牵牛为夫、织女为妇，织女牵牛之星各处河鼓之旁，七月七日乃得一会。”这时“牛郎织女”的故事和七夕相会的情节，已经初具形态了。南北朝时期的《述异记》里有这么一段：“大河之东，有美女丽人，乃天帝之子，机杼女工，年年劳役，织成云雾绢缣之衣，辛苦殊无欢悦，容貌不暇整理，天帝怜其独处，嫁与河西牵牛为妻，自此即废织经之功，贪欢不归。帝怒，责归河东，一年一度相会。”直到民间传说中又对其加以发展。

这个故事是这样的：传说在很久以前，南阳城西的牛家庄有一个叫牛郎的孤儿，随哥哥嫂子生活。嫂子对他不好，给了他九头牛，却让他领十头回来，否则永远不要回去。沮丧之时，他得到高人指点，在伏牛山发现了一头生病的老黄牛。他悉心照料后，才得知老牛原来是天上的金牛星被打下凡间，牛郎成功将其领回家。后来在老牛的指点下，牛郎找到了下凡仙女们洗澡游玩的地方，拿起了其中一位仙女的衣服。那个仙女名字叫“织女”。两人相识后，坠入爱河，后生育有龙凤胎。但此事被王母娘娘发现后，织女被带回了天界。老牛告诉牛郎，它死之后把皮做成鞋就可以腾云驾雾。后来牛郎终于上了天界，眼看就要和织女团聚，却被王母娘娘头上银簪所变的银河拦住去路。天上的喜鹊被他们的爱情感动了，化作“鹊桥”让牛郎织女终于团聚。王母娘娘有些动容，后来允许每年农历七月初七，两人才可在鹊桥相会。之后，每年“七夕”牛郎就把两个小孩放在扁担中，上天与织女团聚，成为一段佳话。

我们认为：童话和神仙故事并不会因物质文明的进步而消亡。它们可以提高孩子们的幻想能力，也可以作为成年人的童年回想，又可以作为各种文化艺术的原料。中国的《牛郎织女》可以和希腊的《奥德

赛》、《金羊毛》，法国的《尼伯龙根指环》等故事并列。2008年，牛郎织女传说入选了国家非物质文化遗产名录。



绘画：张京

嫦娥飞天的神话在《淮南子·外八篇》中记载如下：

“昔者，羿狩猎山中，遇姮娥于月桂树下。遂以月桂为证，成天作之合。”



“逮至尧之时，十日并出。焦禾稼，杀草木，而民无所食。猰貐、凿齿、九婴、大风、封豨、修蛇皆为民害。尧乃使羿诛凿齿于畴华之野，杀九婴于凶水之上，缴大风于青丘之泽，上射十日而下杀猰貐，断修蛇于洞庭，擒封豨于桑林。万民皆喜，置尧以为天子。”

牛郎织女

“羿请不死之药于西王母，托与姮娥。逢蒙往而窃之，窃之不成，欲加害姮娥。娥无以为计，吞不死药以升天。然不忍离羿而去，滞留月宫。广寒寂寥，怅然有丧，无以继之，遂催吴刚伐桂，玉兔捣药，欲配飞升之药，重回人间焉。”

“羿闻娥奔月而去，痛不欲生。月母感念其诚，允娥于月圆之日与羿会于月桂之下。民间有闻其窃窃私语者众焉。”

译文为：

远古时候，天上有十日同时出现，晒得庄稼枯死、民不聊生，一个名叫后羿的英雄，力大无穷，他同情受苦的百姓，于是便拉开神弓，一气射下九个太阳，并严令最后一个太阳按时起落，为民造福。后羿的妻子名叫嫦娥。后羿除了传艺狩猎外，终日和妻子在一起。有不少志士慕

名前来投师学艺，心术不正的逢蒙也混了进来。

一天，后羿到昆仑山访友求道，向王母求得一包不死药。据说，服下此药，能即刻升天成仙。然而，后羿舍不得撇下妻子，暂时把不死药交给嫦娥珍藏。嫦娥将药藏进梳妆台的百宝匣。三天后，后羿率众徒外出狩猎，心怀鬼胎的逢蒙假装生病，没有外出。待后羿率众人走后不久，逢蒙持剑闯入内宅后院，威逼嫦娥交出不死药。嫦娥知道自己不是逢蒙的对手，危急之时她转身打开百宝匣，拿出不死药一口吞了下去。嫦娥吞下药，身子立时飘离地面，冲出窗口，向天上飞去。由于嫦娥牵挂着丈夫，便飞落到离人间最近的月亮上成了仙。

傍晚，后羿回到家，侍女们哭诉了白天发生的事。后羿既惊又怒，抽剑去杀恶徒，逢蒙早已逃走。后羿气得捶胸顿足、悲痛欲绝，仰望着夜空呼唤嫦娥。这时他发现，今天的月亮格外皎洁明亮，而且有个晃动的身影酷似嫦娥。后羿思念妻子，便派人到嫦娥喜爱的后花园里，摆上香案，放上嫦娥平时爱吃的蜜食鲜果，遥祭在月宫里的嫦娥。百姓们闻知嫦娥奔月成仙的消息后，纷纷在月下摆设香案，向善良的嫦娥祈求吉祥平安。从此，中秋节拜月的风俗便在民间传开了。



嫦娥奔月图

水调歌头

明月几时有？把酒问青天。
不知天上宫阙，今夕是何年？
我欲乘风归去，又恐琼楼玉宇，高处不胜寒。
起舞弄清影，何似在人间？
转朱阁，低绮户，照无眠。
不应有恨，何事长向别时圆？
人有悲欢离合，月有阴晴圆缺，此事古难全。
但愿人长久，千里共婵娟。

宋代词人苏轼的一首《水调歌头》更是把自己与月中的飞天嫦娥联系起来，意境高远。至于《西游记》、《封神榜》等文学作品中那些在天地之间自由往来的神怪们，同样令人赞叹。他们对现代航天登月飞天、星际探测，也许会有一些思想启迪视野开放作用。

2.1.4 天文星图、历法和观测仪器

东汉中期（公元2世纪）张衡撰写了关于天文方面的著作《灵宪》。

张衡（公元78—139年），字平子，汉族，南阳西鄂（今河南南阳市石桥镇）人，我国东汉时期杰出的天文学家、数学家、发明家、地理学家、制图学家、文学家、学者。在汉朝官至尚书，为我国天文学、机械技术、地震学的发展作出很大的贡献。

张衡是东汉中期浑天说的代表人物之一。他指出月球本身并不发光，月光其实是日光的反射；他还正确地解释了月食的成因，并且认识到宇宙的无限性和行星运动的快慢，与距离地球远近的关系。

张衡观测记录了2500颗恒星，创制了世界上第一架能比较准确地表演天象的漏水转浑天仪，还制造出了指南车、自动记里鼓车、飞行数里的木鸟等。

张衡共著有科学、哲学和文学著作32篇，其中天文著作有《灵宪》和《灵宪图》等。

为了纪念张衡的功绩，人们将月球背面的一个环形山命名为“张衡山”，将1802号小行星命名为“张衡星”。



张衡

20世纪中国著名文学家、历史学家郭沫若对张衡的评价是：“如此全面发展之人物，在世界史中亦所罕见，万祀千龄，令人敬仰。”

《晋书·天文志》中还记载了吴国太史令陈卓编制出了包含283个星官，1464颗星的星座体系。

现存最古老的星图是敦煌星图，约绘于唐中宗时期，包含星星1350颗。

最完整的古星图是宋淳祐七制所刻的石刻天文图，现存于苏州。包含有1440颗星，还有坐标和二十八宿分界，标出了赤道和黄道，刻出了银河。

对行星观测方面，长沙马王堆出土《五星占》，成书在公元前170

年左右，记载了从公元前246年至公元前177年的行星位置和周期。行星的位置和运动为历代历法重要内容。

汉代司马迁的《史记·天宫书》记载了行星逆行的发现。

下面介绍杰出的司马迁。

司马迁（公元前145—前90年），字子长，西汉夏阳（今陕西韩城南）人，中国西汉伟大的史学家、文学家、思想家、天文学家，司马谈之子。他曾任太史令，因李陵之事有所辩解，下狱，受腐刑，后任中书令时，发奋继续完成所著史籍，被后世尊称为“史迁”、“太史公”、“历史之父”。

他以其“究天人之际，通古今之变，成一家之言”的史识创作了中国第一部纪传体通史《史记》（原名《太史公书》或《太史公记》）。这本史书被公认为是中国史书的典范，记载了从上古传说中的黄帝时期，到汉武帝元狩元年，长达3000多年的历史，是“二十五史”之首，被鲁迅誉为“史家之绝唱，无韵之《离骚》”。



司马迁从小就受到良好的教育。

司马迁

在《史记·太史公自序》上，他说：“迁生龙门，耕牧河山之阳。年十岁则诵古文。”二十岁时，他从长安出发，足迹遍及江淮流域和中原地区，所到之处考察风俗，采集传说。“（他）二十而南游江、淮，上会稽，探禹穴，窥九疑，浮沅、

湘。北涉汶、泗，讲业齐鲁之都，观夫子遗风，乡射邹峰；厄困蕃、薛、彭城，过梁、楚以归。于是迁仕为郎中，奉使西征巴、蜀以南，略邛、笮、昆明，还报命。”二十五岁时，他又以使者监军的身份，出使西南夷，担负起在西南设郡的任务，足迹遍及“邛、笮、昆明”等地。

汉武帝元封元年（公元前110年），他的父亲司马谈去世。三年之后，司马迁承袭父职，任太史令，同时也继承了父亲遗志（司马谈临终曾对司马迁说：“余死，汝必为太史；为太史，无忘吾所欲论著矣。”），准备撰写一部通史。汉武帝太初元年（公元前104年），司马迁与唐都、落下闳等共同订立了“太初历”，该历法改变了秦代使用的颛顼历以十月为岁首的习惯，而改以正月为岁首，从而为中国的农耕社会奠定了其后两千年来所尊奉的历法基础。之后，司马迁便潜心修史，开始了《史记》的写作。公元前98年，因李陵案获罪下狱。李陵被灭族，而司马迁为著作史记而忍辱苟活，自请宫刑。

司马迁的著作，除《史记》外，《汉书·艺文志》还著录赋八篇，均已散失，唯《艺文类聚》卷30中引征《悲士不遇赋》的片段和有名的《报任安书》（即《报任少卿书》）遗存于世。《报任安书》表白了他为了完成自己的著述而决心忍辱含垢的痛苦心情，是研究司马迁生平思想的重要资料，也是一篇饱含感情的杰出散文。《悲士不遇赋》也是他晚年的作品，抒发了受腐刑后和不甘于“没世无闻”的愤激情绪。

史记：公元前91年（征和二年），《史记》全书完成。全书130篇，526 500余字，包括十二本纪、三十世家、七十列传、十表、八书，对后世的影响极为巨大，被称为“实录、信史”。列为前“四史”之首，与《资治通鉴》并称为史学“双璧”。

至治思想：司马迁借老子之口说出了被他改造过和发展了的至治理

想。保留了道法自然的内核，有意识地剔除了反映原始社会“小国寡人”和奴隶社会“使民”如何的思想，体现了深得道家精髓的“与时迁移”思想。其核心是天道自然，要旨是人民的足欲。

天文成就：两汉时期的天文星象家中，首先要提到司马迁。大家都知道司马迁是一个伟大的史学家，而不知他也是一位在天文星象方面造诣很深的专家。其实只要仔细读其《史记》的《天官书》、《律书》、《历书》，就可明白称他为天文星象专家绝非虚誉。

司马迁继承父亲的遗志，完成“推古天变”之任务，并明确表述为“通古今之变，究天人之际”，其结论表述在《天官书》中，即据春秋242年之间日食三十六、彗星三见等星象，联系点天子衰微、诸侯力政、五伯代兴及到战国及秦汉之际的社会变乱动荡，而总结出天运三十年一小变，一百年一中变，五百年一大变，三大变为一纪，三纪而大备的“大数”，最后才认为“天人之际续备”。这是司马迁天文学之应用的最重要之范例，在整个星学历史上占有最高地位。

此外，司马迁又总结了战国以来的天文学的基本原则，大意为：“我仔细检查史书的记载，考察历史上的事变，发现在100年之中，五星皆有逆行现象。五星在逆行时，往往变得特别明亮。日月的蚀食及其向南、向北的运行，都有一定的速度和周期，这是星象学所要依据的最基本的数据。而星空中的紫宫星垣和东西南北四宫星宿及其所属的众多星辰，都是位置不变的，它们的大小程度和相互间的距离也是一定不变的，它们的分布排列象征着天上五官的位置。这是星象学中作为‘经’，亦即不变的依据。而水、火、金、木、土星则是上天的五位辅助，它们的出现隐伏也有一定的时间和周期，但其运行速度快慢不均。这是天文学中的‘纬’，亦即经常变动的部分。把这些固定的和变动的两种星象结合起来，就可以预测人事的变化了。”

历史评价：

（1）司马迁承继其父司马谈的黄老之学，坚持朴素唯物主义观点，对神秘化的阴阳五行学说采取讥评的态度。他还以造诣甚深的自然科学知识（天文历法）为依据，认为阴阳、四时和二十四节气等是不可更改的，“春生夏长，秋收冬藏”是自然界的客观规律，人们必须遵守，不可违背。由此，他对当时占统治地位的“天人感应”的神秘主义学说进行了有力的讽刺和批判。

（2）司马迁通过对当时社会现象的观察和以历史上的大量事实为根据，说明善人往往不得善报而遇灾受害者“不可胜数”，但恶人却有“终身逸乐，富贵累世不绝”、“竟以寿终”的。揭露了所谓“天道有知”、“天之报施善人”的欺人之谈。

（3）反对神鬼和求仙可致长生不死的迷信。

（4）司马迁试图从人们生活的物质基础即经济生活中找寻社会现象和社会意识问题的根据。他把人们从事农、矿、工、商等经济活动的历史看成和自然现象的变化一样是有规律可循的；并认为社会生产的分工和发展是被人们物质生活的需要所推动的，而不是什么政治力量和神的意志所能创造的。

（5）司马迁认为社会上等级和奴役关系的产生，起源于彼此占有财富的悬殊，他揭示出富者剥削和奴役贫者是人类社会的现象，并非“天命”的安排。

（6）司马迁继承了《管子》的道德观，认为财富的占有情况也决定人们的道德观念。司马迁抨击“天道”，肯定“利”和“欲”，认为有“利”即有“德”。

(7) 司马迁具有较进步的道德观。他充分发扬前人的以人为本思想，对于当世、历史、人民、国家有所贡献和作为的，尤其是不符合世俗的道德观念，予以记述和歌颂，这与当时的封建伦常、道德观念大相径庭。

(8) 司马迁主张从不断变化发展的观点去考察国家成败兴亡的道理。他以“通古今之变”的历史进化观点来观察历史。他在《自序》中所说的“原始察终，见盛观衰”，正是他“通古今之变”的方法论。

这一时期的观测仪器：

(1) 测日仪器：圭表。

(2) 测时仪器：日晷、铜壶滴漏。

(3) 测星仪器：浑仪，最早西汉时期（公元前1世纪）制造；简仪，元代郭守敬对浑仪进行了根本变革，创制简仪。

(4) 天球仪：最早西汉时期制造“浑象”；东汉张衡制造“水运浑象”；唐朝时期制造“水运浑天”，有报时机构。

历法：

(1) “古六历”（黄帝历、颛顼历、夏历、殷历、周历、鲁历）：先秦至汉初使用。取平均年长365.25日；朔望月长29.530 851日；采用18年置7闰。颛顼历取立春为岁首，其余取冬至为岁首。

(2) 太初历：汉武帝元封七年（公元前104年）制订。首次规定没有中气的月份为闰月；把二十四节气订入历法。第一次计算了日月食发生的周期。

(3) 大明历：南北朝时期刘宋孝武帝大明六年（公元462年）成历，由祖冲之制订。取年长365.2428日；这是第一部计及岁差的历法，采用391年加144年闰的闰周。但这一历法直到公元510年才采用，使用达80年。

祖冲之（公元429—500年）是我国杰出的数学家、科学家。南北朝时期人，汉族，字文远。祖籍河北涿水，出生地南京。其主要贡献在数学、天文历法和机械三方面。

天文历法：在古代，我国历法家一向把19年定为计算闰年的单位，称为“一章”，在每一章里有7个闰年。也就是说，在19个年头中，要有7个年头是13个月。这种闰法一直采用了一千多年。



祖冲之

祖冲之吸取了前人的先进理论，加上他自己的观察，认为19年7闰的闰数过多，每200年就要差1天。因此，他提出了391年内144闰的闰周法。这个闰法在当时算是最精密的了。

除了改革闰法以外，祖冲之在历法研究上的另一重大成就，是破天荒第一次应用了“岁差”。

根据物理学原理，刚体在旋转运动时，假如丝毫不受外力的影响，旋转的方向和速度应该是一致的；如果受了外力影响，它的旋转速度就

要发生周期性的变化。地球就是一个表面凹凸不平、形状不规则的刚体，在运行时常受其他星球引力的影响，因而旋转的速度总要发生一些周期性的变化，不可能是绝对均匀一致的。因此，每年地球绕太阳运行一周，不可能完全回到上一年的冬至点上，总要相差一个微小距离。按现代天文学家的精确计算，大约每年相差50.2秒，每71年8个月向后移1度。这种现象叫作岁差。

随着天文学的逐渐发展，我国古代科学家们渐渐发现了岁差的现象。汉代多位学者都曾观测出冬至点后移的现象，不过他们都还没有明确地指出岁差的存在。到东晋初年，天文学家才开始肯定岁差现象的存在，并且首先主张在历法中引入岁差，算出冬至日每50年退后1度。



我国古代天文仪器

后来到南朝时，祖冲之继承了前人的科学研究成果，不但证实了岁差现象的存在，算出岁差是每45年11个月后退1度，而且在他创制的《大明历》中应用了岁差。但是因为他所根据的天文史料都还是不够准确的，所以他提出的数据自然也不可能十分准确。尽管如此，祖冲之把岁差应用到历法中，在天文历法史上仍是一个创举，为我国历法的改进揭开了新的一页。到了隋朝以后，岁差已为很多历法家所重视了，像隋朝的《大业历》、《皇极历》中都应用了岁差。

祖冲之在历法研究方面的第三个巨大贡献，就是能够求出历法中通常称为“交点月”的日数。所谓交点月，就是月亮连续两次经过“黄道”和“白道”的交叉点时，前后相隔的时间。黄道是指地球人看到的太阳运行的轨道，白道是地球人看到的月亮运行的轨道。交点月的日数是

可以推算得出来的。祖冲之测得的交点月的日数是27.212 23日，比过去天文学家测得的要精密得多，同近代天文学家所测得的交点月的日数27.212 22日已极为近似。在当时的天文学水平下，祖冲之能得到这样精密的数字，成绩实在惊人。

由于日食和月食都是在黄白交点附近发生，所以推算出交点月的日数以后，就更能准确地推算出日食或月食发生的时间。祖冲之在他制订的《大明历》中，应用交点月推算出来的日、月食时间比过去准确，和实际出现日、月食的时间都很接近。

此外，祖冲之对木、水、火、金、土等五大行星在天空运行的轨道和运行一周所需的时间，也进行了观测和推算。我国古代科学家算出木星（古代称为“岁星”）每12年运转1周。西汉刘歆作《三统历》时，发现木星运转1周不足12年。祖冲之更进一步，算出木星运转1周的时间为11.858年。现代科学家推算木星运行的周期约为11.862年。祖冲之算得的结果，同这个数字仅仅相差0.04年。此外，祖冲之算出水星运转1周的时间为115.88日，这同近代天文学家测定的数字在两位小数以内完全一致。他算出金星运转1周的时间为583.93日，同现代科学家测定的数字仅差0.01日。

祖冲之根据上述的研究结果，成功创制了当时最科学、最进步的历法——《大明历》。这是祖冲之科学研究的天才结晶，也是他在天文历法上最卓越的贡献。祖冲之在天文历法方面的成就，大都包含在他所编制的《大明历》及为大明历所写的驳议中。

公元462年，祖冲之把精心编成的《大明历》送给政府，请求公布实行。根据《大明历》来推算元嘉十三年（公元436年）、十四年（公元437年）、二十八年（公元451年）、大明三年（公元459年）的四次

月食都很准确，但用旧历法推算的结果误差就很大。

但是，由于权贵保守派的阻挠，一直到梁朝天监九年（公元510年），新历才被正式采用，可是那时祖冲之已去世10年了。

精算圆周率：

《隋书·律历志》留下一段关于圆周率（ π ）的记载，祖冲之算出 π 的真值在3.141 592 6和3.141 592 7之间，相当于精确到小数第7位，简化成3.141 592 6，成为当时世界上最先进的成就。祖冲之入选世界纪录协会世界第一位将圆周率值计算到小数第7位的科学家，创造了中国纪录协会的世界之最。这一纪录直到15世纪才由阿拉伯数学家打破。

祖冲之还给出 π 的两个分数形式：22/7（约率）和355/113（密率），其中密率精确到小数第7位，这在西方直到16世纪才由荷兰数学家奥托重新发现。祖冲之还和儿子一起圆满地利用“牟合方盖”解决了球体积的计算问题，得到了正确的球体积公式。

求算圆周率的值是数学中一个非常重要，也是非常困难的研究课题。中国古代许多数学家都致力于圆周率的计算，而公元5世纪祖冲之所取得的成就可以说是圆周率计算的一个跃进。祖冲之经过刻苦钻研，继承和发展了各位数学家、科学家前辈的优秀成果。他对于圆周率的研究结果，是对我国乃至世界的一个突出贡献。他的精确推算值被命名为“祖冲之圆周率”，简称“祖率”。

制造机械：

指南车是一种用来指示方向的车子。车中装有机件，车上装有木人。车子开行之前，先把木人的手指向南方，不论车子怎样转弯，木人

的手始终指向南方不变。这种车子结构已经失传，但是根据文献记载，可以知道它是利用齿轮互相带动的结构制成的。

祖冲之仿造前人所制的指南车内部机件是全铜的。它的构造精巧、运转灵活，无论怎样转弯，木人的手常常指向南方。

祖冲之也制造了很有用的劳动工具。古代劳动人民很早就发明了利用水力舂米的水碓和磨粉的水磨。祖冲之在前人基础上进一步加以改进，把水碓和水磨结合起来，生产效率就更加提高了。这种粮食加工工具，称为“水碓磨”，现在我国南方有些农村还在使用着。

祖冲之还设计制造过一种千里船，可能是利用轮子激水前进的原理造成的，一天能行500多千米。

祖冲之还根据春秋时代文献的记载，制作了一个“欹器”。欹器是古人用来警诫自满的器具。器内没有水的时候，是侧向一边的。里面盛水以后，如果水量适中，它就竖立起来；如果水满了，它又会倒向一边，把水泼出去。这种器具，晋朝的学者曾试制三次，都没有成功；祖冲之却仿制成功了。由此可见，祖冲之对各种机械都有深刻的研究。

撰写缀术：

祖冲之还曾写过《缀术》五卷，这是一部内容极为精彩的数学书，很受人们重视。《缀术》在唐代被收入《算经十书》，成为唐代国子监算学课本。唐朝的官办学校的算学科中规定：学员要学《缀术》四年；政府举行数学考试时，多从《缀术》中出题。《缀术》一书，汇集了祖冲之父子的数学研究成果。这本书内容深奥，以至“学官莫能究其深奥，故废而不理”。可见《缀术》的艰深。《缀术》曾经传至朝鲜，但到北宋时就已遗失了。

祖冲之的成就不仅限于自然科学方面，他还精通乐理，对于音律很有研究。

他又著有《易义》、《老子义》、《庄子义》、《释论语》等关于哲学的书籍，但可惜的是，现在都已经失传了。

历史评价：

祖冲之在天文、历法、数学以及机械制造等方面的辉煌成就，充分表现了我国古代科学的高水平。首先，当时社会生产正在逐步发展，需要有一定的科学归纳、总结来配合前进，科学的进步具有了较好的基础和推动力。祖冲之正逢其时，取得了多个学科专业方面的成绩。其次，从上古到这时候，在千百年的实际活动中，已经积累了不少科学技术成果，祖冲之就在前人已搭建的舞台上，以自己卓越的学识实现了创新提高，做出了他的成绩。至于祖冲之个人的禀赋，认真学习、刻苦钻研，不迷信古人，不畏惧守旧势力，不避艰难，自然也都是取得杰出成就的重要原因。

祖冲之不仅是我国历史上杰出的科学家，在世界科学发展史上也有崇高的地位。我们应该纪念像祖冲之这样的科学家，珍视他们的宝贵遗产。为纪念这位伟大的古代科学家，人们将月球背面的一座环形山命名为“祖冲之山”，把1888号小行星命名为“祖冲之小行星”。

2.1.5 小结

这一时期中国的宇宙及天文学得到大发展，获得诸多新成果，如精确的历法编算、精致的天文仪器、富有想象力的飞天神话和先进的宇宙观念等方面都取得新成就。以屈原、荀子、张衡、司马迁和祖冲之等一

批杰出的学者、大家为代表，继承了中华民族千百年历史积淀的财富，通过个人天赋创造及艰辛的劳动，开创出了一个繁荣兴盛的好局面。但是，由于政治制度的专制和权贵保守势力的专横，一直在阻碍着科学技术和新思想的发展，这必然隐含着衰落的危机。

2.2 埃及、美索不达米亚、印度的宇宙及天文学进展

2.2.1 埃及

埃及的民用历沿用先前的太阴历，以天狼星偕日升（约合公历的7月19日前后）为一年的起始，这最早见于公元前3世纪第一王朝的历史记载，但实际采用时间则要更早些。天狼星偕日升标志着尼罗河在旧王国的国都——孟菲斯泛滥。

由于365日比实际的回归年长度要短 $1/4$ 天左右，即每隔4年，新年时间就要比实际提前1天，每过1460个回归年则提前1年，这就是所谓的“天狼星周期”。有说法认为古埃及历年与回归年之间的差异正是“徘徊年”之名的由来。

这时，法老颁布命令，决定每4年设置1个闰年。不过这一改革当时遭到了农民的反反对，原因是旧历与农业生产紧密关联。直到公元前1世纪，置闰规则才为改革后的亚历山大历真正采用，并于公元前22年首次置闰，闰日不属于任何一个月，而是附加在空余出的几天中。

准太阴历的基础是民用历法，但并不取决于天狼星，主要作用是确定宗教节日的时间。其置闰法则是，倘若太阴年的新年早于民用年，则设置闰月，后来则有了25年的置闰法则。

此外古埃及还有类似现代财政年度的设置，财政年始于 II Peret（埃及古历冬季）第1天，终于 I Peret第30天，其他同民用历法。在一些天文记载中，还出现了若干不用法老年号表示的特定时期，很是

烦琐。

后来衍生出的公历1年设置为365天，是古埃及历法的一大贡献。另一贡献是对黄道十二宫的划分。他们在新王国时期已经知道了40多个星座，考古学家在墓地和神庙中获得了类似“星位图”的记录。

埃及的拜神主义通常将东地平线作为他们精神源泉的起点。这种想法也深深影响了埃及人制定历法的方式。如同美索不达米亚人，埃及人最初也使用基于月亮运行周期而制定的历法，但是区别是：美索不达米亚人以日落时西方的新月作为每月的开始，而埃及人却以日出时东方的新月作为每月的开始。当西方文艺复兴的影响波及埃及的时候，埃及人的信念开始动摇了。他们很快意识到，虽然月亮历在大多数情况下有实用价值，但它却有许多缺陷。其中最大的缺点就是每当12个月月亮历月循环一次时，每年都会多出11天。为了使历法能够永远符合节气，只能每隔两年附加上1个月来弥补。

埃及人曾经试图废除传统月亮历，而引进基于太阳运行周期的新“国民”历法。新历法用于为国家权力机构管理国家服务，但月亮历法仍然保留于僧侣事务与日常公众活动。通过记录与太阳同时出没的天狼星的运行情况，他们甚至已经能计算出太阳绕黄道1周的时间为365天。为方便起见，他们简单地把1年划分为12个月，每月30天。另外的5天单独作为一个时段，这段时间他们主要用于欢庆。每个月又被划分为3周，每周10天。每周的开始用在黄昏时分升起的特殊恒星或星团来确定，而不是努力地契合月相的变化。为了使新历法更有效，他们还把天球细化为36个部分，这样夜晚的时间就可以通过恒星的起落来确定。

埃及人的新历法把一年分为36个区间，每个区间又由一颗专门的具有神性的恒星掌控。这就产生了占星学理论体系中所谓“古典占星学10°

分区”的概念，虽然在此以后这个概念常常被认为是中世纪占星学理论的专利。这个概念将原本已经被划分为30°的黄道带又被细化为10°分区，并且每个分区又由一颗行星掌控。因此恒星时变成了神的时间——也就是人类命运的尺度。这个事实虽然直到象征主义被取代时才确切地知道，但在公元前200年建立的神庙已经向人们展示了10°分区在占星学上的运用。

埃及人神化时间概念的详情是值得人们关注的。在埃及人的信仰中，每1个恒星都被描绘为1位神，并掌管黄道上的一个10°区间。它们被确定的方法是，每周开始时最先升起的那颗恒星就是这1周的统治者也就是这1周的神。在早期，恒星钟被制造出来以确定每一周的统治恒星。这个简略的装置甚至在夜里也能提示每一个10°区间到来的准确时间。因此星相家们就可以通过记录这些时间，用图表计算，标出夜晚的每个小时了。但实际上，现存的12个在埃及神庙中被发现的事例，已经证明这种方法在当时更为被看重为是一种为亡灵在阴间提供财富的重要方式。神庙中还有一个用恒星在天空中运行位置排列成的坐着的人形图案。用这些恒星位置就可以很容易地表示出1年中14天为1周期的晚上的准确时间。

古埃及被神化了的时间历法的更进一步运用，是能通过恒星与星群的运动指明任意时刻的掌控恒星。这些作用在太阳领域下白天12小时里的守护神在神话中显得特别重要。当然，夜晚里的守护神同样也是重要的。当太阳神落到地平线以下时，他就必须在黑暗世界里通过每1小时区各个神的领域。这些神为太阳神打开大门，并把他送往下一个领域，而开门的密码只有这些暗夜守护神们自己知道。小时区间神的概念同样被运用到占星学中，特别当占星上的判断需要精确到某个或某几个小时时。但后来行星被统一定为每个小时的守护神，并成为占星询问中非常

有效的描述性征兆。

但行星时不像通常的时钟那样固定而连续的，它在长度上是会变动的，这也是追随埃及历法的一种表现。通过水钟的发明，埃及人第一次创造了一天24小时制，包括白天12小时，夜晚12小时。在春分、秋分时，白天、黑夜的24个小时是等量平分的，但在一年里的其他时间，每个小时的长度就是常常变化的了。它们的计算方法是：把日出到日落划分为白天12小时，再把日落到日出划分为黑夜12小时。所以白天的12小时在夏天就比较长，而在冬天就比较短；黑夜的12小时在夏天就比较短，而在冬天就比较长。

自从日落日出被分别定义为太阳的“死亡”与“重生”时，它们就具有非常重大的意义。太阳的出现或消失，改变了人的整个生命模式，从活跃到寂静，埃及人为我们留下的时间历法对于标明这样的转变显得非常宝贵。宗教仪式往往在日出时举行（比如英国教堂的晨祷，天主教子夜或黎明的祈祷）——牧师的职能之一——调用白天行星的神性能力。这个行星被看待为此时天上的统帅，同时这个小时里的守护行星也是这个人此生的代理者。占星学上每小时统治次序以行星占星术顺序排列：月亮、土星、木星、火星、太阳、金星、水星。另一种宗教仪式（晚祷即天主教每天七段祈祷中的晚课）在每天第八小时举行，这时正好此小时的统治者变回为统治白天的那个行星。虽然我们现在一周七天的名字仍以这些行星统治者来命名，但行星时却被认为只适合在神秘主义学说中运用。24小时等长的制度在希腊文化时代被引入并发展至今，但这种时间制度彻底分离了每天与季节的联系。

2.2.2 美索不达米亚

巴比伦人以新月初见为一个月的开始，这个现象发生在日月合朔后一日或二日，决定于日月运行的速度和月亮在地平线上的高度。为了解决这个问题，天文学家自公元前311年开始制定日、月运行表，这个表只有数据，没有任何说明。它的奥秘在19世纪末终于被现代学者揭开。他们发现，第4栏是当月太阳在黄道十二宫的位置，第3栏是合朔时太阳在该宫的度数（每宫从 $0^{\circ}\sim 30^{\circ}$ ），第3栏相邻两行相减即得第2栏数据，它是当月太阳运行的度数。若以月份为横坐标，以太阳每月运行的度数为纵坐标绘图，便可得3条直线。前3点形成的直线斜率为 $+18'$ ，中间6点形成的直线斜率为 $-18'$ ，后4点形成的直线斜率为 $+18'$ 。若就连续若干年的数据画图，就可得到一条折线。在这条折线上两相邻峰之间的距离就是以朔望月表示的回归年长度，1回归年=12.5朔望月。

巴比伦天文表：在这种日月运行表中，一些项目有18栏之多。除上述4栏外，还有昼夜长度、月行速度变化、朔望月长度、连续合朔日期、黄道对地平的交角、月亮的纬度，等等。有日月运行表以后，计算月食就很容易了。事实上，远在约公元前9世纪时，人们就已经知道：月食必发生在望，而且只有当月亮靠近黄白交点时才行。

准确的周期测算：巴比伦人不但对太阳和月亮的运行周期测得很准确，朔望月的误差只有0.4秒，近点月的误差只有3.6秒，对五大行星的会合周期也测得很准确。

2.2.3 印度

印度人很早就开始了天文历法的研究。在吠陀时代，他们已有不少天文历法知识。那时，他们把一年定为360日，分为12个月，也有置闰的方法。我国唐朝时，古印度学者后裔瞿昙悉达著有《天元占经》一

书。这部书里所介绍的“九执历”是那时印度较先进的历法。

这部历法规定，1恒星年为365.2726日（今测值为365.256 36日），1朔望日为29.530 583日（今测值为29.530 589日），采用了19年7闰的置闰方法。

古印度比较著名的天文历史著作，是公元前6世纪的《太阳悉檀多》。这部著作讲述了时间的测量、分至点、日月食、行星的运动和测量仪器等许多问题。

这部书成为古印度天文学家著作的范本，它同时还是古印度最重要的数学著作之一，对古印度天文学和数学有很大的影响。

古印度还有一部杰出的天文学著作，是公元前5世纪后期圣使所著的《圣使集》。其中提到天球运动是地球绕地轴旋转而见到的现象，但这一超时代的正确见解，并没有被当时的人接受。在这部天文学著作中，还讨论了日、月和行星的运动，以及推算日月食的方法等。

公元505年，古印度就有了综合性的天文学著作《五大历数全书》。此书是一部汇集了古印度5种最重要的天文学史的著作。这部书在天文学史上很有参考价值，作者虽没有什么自己的见解，但却把前人的成果阐述得很系统、很清晰。

古印度人在天文历法方面虽然做了许多有意义的工作，但是他们不十分注重实际的天文观测，因此在长时间内都还只有一些比较简单的观测仪器，直至18世纪才在德里等地建立起一些有较为复杂的观测仪器的天文台。

在古印度，不同时代的人对宇宙有着不同的看法。如在吠陀时代，

人们认为天地的中央是一座名叫“须弥山”的大山，日、月都绕此山运行，太阳绕行1周即为1昼夜。而《太阳悉檀多》则认为大地是球形，北极是山顶，此山名叫“墨路山”，那是神的住所，日、月和五星的运行是一股宇宙风所驱使，一股更大的宇宙风使所有天体一起旋转。此外，印度著名的天文学家作明（公元1114—？）在他的《历数全书头珠》的著作中，主张地球是靠自身的力量固定于宇宙之中，其上有七重气，分别推动日、月和五星的运行。这时，作明的想法已受到了古希腊人的影响。

2.2.4 小结

由于航海技术的发展，人员、贸易往来增多，学者们也频繁游学、往返于各地，使得这个时期埃及、美索不达米亚、印度与希腊、罗马的宇宙及天文学逐步融汇到一起，而以希腊、罗马的理论学说作为代表，广为流传于后世了。

2.3 希腊、罗马的宇宙及天文学进展

这一时期，希腊、罗马的诸派学者们对于宇宙的结构、构型继续展开激烈地争斗，其中最关键的问题是：宇宙的中心是地球还是太阳？地球是静止的还是运动着的？概括起来，主要有“日心说”和“地心说”等学派。然而，由于生产、科技和社会发展水平的局限，特别是后来教会统治者的需要，地心说占据着正宗、统治地位。

2.3.1 中心火焰说

古希腊爱奥尼亚学派的泰勒斯认为地球是浮在水上的一个扁平圆盘。他的学生第一个明确提出地球是一个球，居于世界的中央。而毕达哥拉斯学派的学者则设想宇宙有一个中心火焰，地球和“对地”这两个天体处与中心火焰的两侧，在不同的同心圆周上运行，太阳则在第三个同心圆周上。但是这种中心火焰说遭到了社会实践的驳诘。当时地中海沿岸各地之间的贸易往来日渐频繁，航海所及范围逐渐扩大。迦太基的航海者由地中海向西作了一次有名的航行，使人们的眼界越过了当时被认为是西方边界的直布罗陀海峡。此外，公元前327年亚历山大大帝东征到印度，又使人们的眼界向东直抵印度河流域。地理的地平圈扩大了。可是，即使在这扩大的地平圈上，仍然窥测不到隐藏在地球背面的“对地”，也看不到中心火焰。这些事实证明了“对地”及中心火焰是虚构的。毕达哥拉斯学派的中心火焰说失败了，但这毕竟是关于地球布局及运行的第一个推测。

2.3.2 日心说的先驱

公元前4世纪的希腊学者曾提出一项大胆的说法：我们所见的天体圆周式运动，可以假定为地球自转造成的，并认为金星和水星运行轨道的中心是太阳，而不是地球。此理论虽并不正确，却打破了地球不动的既有观念，但当时仍不为人所信，“因为大地稳稳当当、一丝不动的直接感觉似乎与此矛盾”。

另一学者阿里斯塔克则更为激进，被后世天文学家誉为“古代世界的哥白尼”。其理论与哥白尼的日心说有着惊人的相似性：他同样主张太阳与恒星是固定的，地球运行于环绕太阳的轨道上，并且处于这轨道的中部，并以自转来解释天体的每日循环，以公转来解释太阳绕黄道的表现路径。“对于应该产生但没有被观测到的恒星周年视差，阿里斯塔克推测地球轨道半径与地球到恒星的距离相比是微不足道的。”

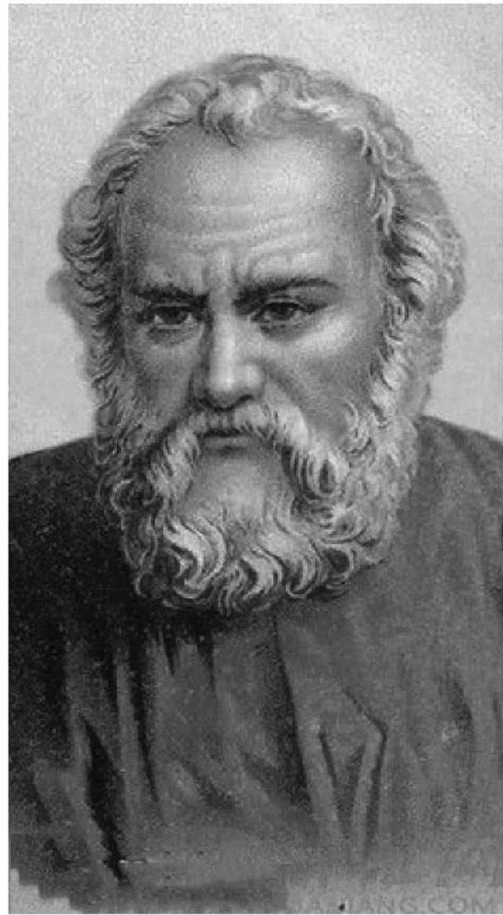
阿里斯塔克的观点太过前卫，且存在“为什么地球上的天体没有抛后”、“恒星周年视差”等依靠当时技术无法解决的问题。不仅仅他的学说没有被当时天文学界接受，甚至本人都被宗教团体诘难。

著名学者阿基米德曾运用水力制作一座天象仪，球面上有日、月、星辰、五大行星。根据记载，这个天象仪不但运行精确，连何时会发生月食、日食都能加以预测。而晚年的阿基米德也开始怀疑地球中心学说，并猜想地球有可能绕太阳转动。

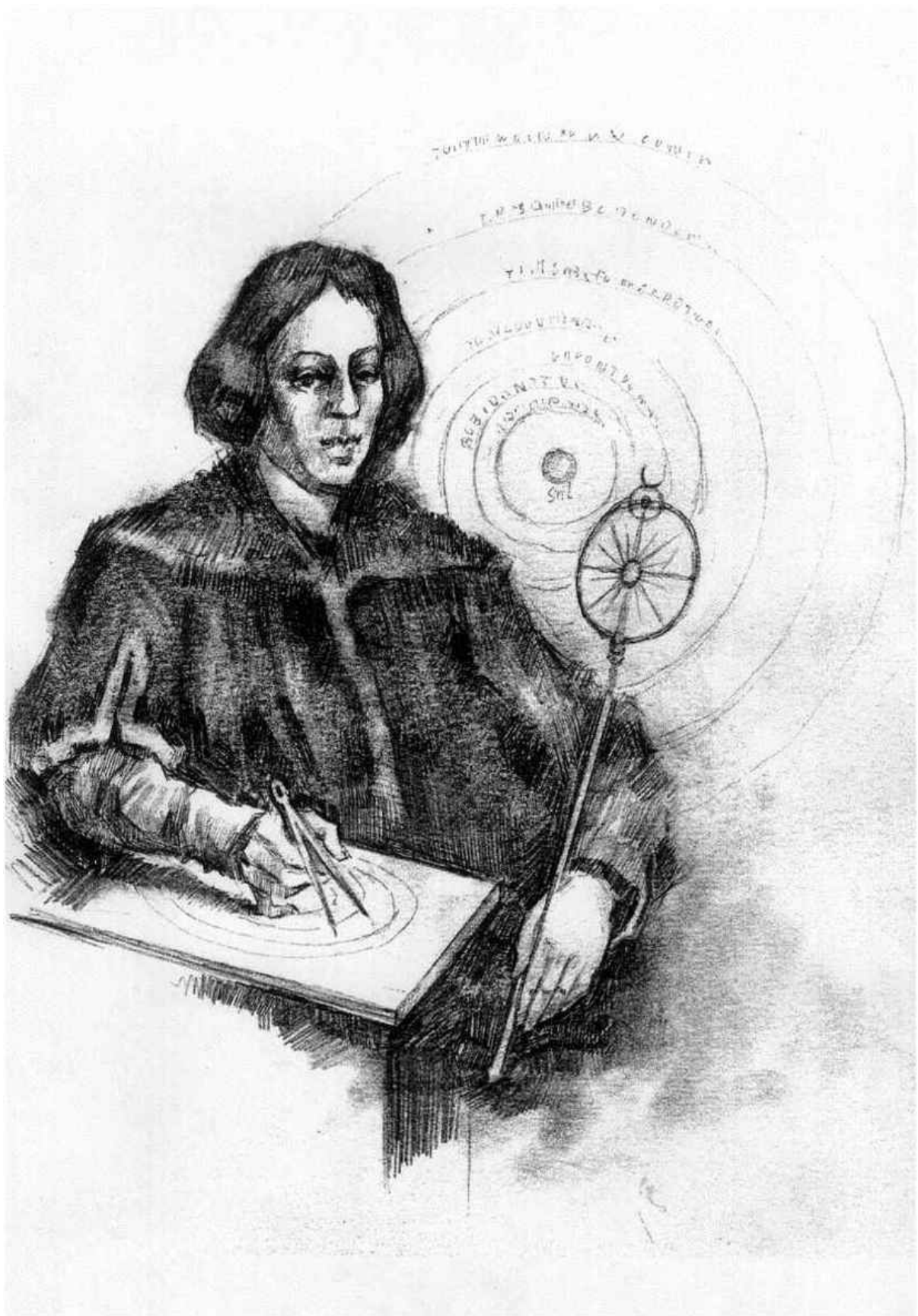
阿基米德是一位非常杰出的科学家，和雅典时期的其他科学家有着明显不同，他既重视科学的严密性、准确性，要求对每一个问题都进行精确、合乎逻辑的验证；又非常重视科学知识的实际应用。他非常重视试验，亲自动手制作各种仪器和机械。他一生设计、制造了许多机构和机器。除了杠杆系统外，值得一提的还有举重滑轮、灌地机、扬水机以及军事上用的抛石机等。被称作“阿基米德螺旋”的扬水机至今仍在埃及

等地使用。他亲身投入保卫家园的战斗，但不幸被愚昧的敌人杀害，令人无限惋惜。他的“利用浮力原理鉴别金王冠是否掺假的故事”、“利用杠杆可以撬动地球的豪言”至今仍在流传。

他沉迷于研究、创造。他在理论研究，尤其是在数学和天文方面有很多杰出的成果。在数学上，他曾利用“逼近法”算出球面积、球体积、抛物线、椭圆面积，后世的数学家依据这样的“逼近法”，并加以发展成为近代的“微积分”。他更研究出螺旋形曲线的性质，现今的“阿基米德螺线”曲线，就是因为纪念他而命名。另外他在《恒河沙数》一书中，创造了一套计大数的方法，简化了计数的方式。



阿基米德



绘画：张京

2.3.3 物性论

在这里，特别要提到卓越的罗马诗人、唯物哲学家卢克莱修（公元前96—公元后55年）所著的以《物性论》为题目的一首长诗。全诗7000余行，以日常所见的事物、通俗易懂的词语、优美引人的诗句阐述了他的宇宙基本原理和一般规律。（中文有多个译本）因此，长诗《物性论》的内容可以看作古希腊、罗马在关于宇宙本源问题上，唯物论者做出的综合性总结。

对于宇宙的基本规律，诗人一开头就以大无畏的无神论者姿态出现在世人面前。诗中这样写道：

恐惧所以能够统治亿万众生，
只是因为人们看见大地环宇
有无数他们不懂其原因的事象，
因此以为有神灵操纵其间。
而当我们一朝知道，
无中不能生有，我们就会
更清楚地猜到我们寻求的，
万物由之造成的那些元素，
以及万物的造成如何未借神助。

“无中不能生有”是唯物论诗人得出的第一宇宙规律。宇宙第二规律是自然能够把一切东西分解为原子，而不能把它消灭、归为乌有。这样就从正反两方面阐明了唯物论的基本宇宙规律，其中朴素地孕育着现代科学所公认的物质不灭定律。诗中写道：

然而事实上，既然每一种事物都从
特定的种子里创造出来，那个它从中
出生并进入光明之岸的源头
就是它的质料和原子。
这就是为什么万物不能从万物中出生，
因为每个物体里都蕴藏着它独特的能力，

.....

但如果穿越那漫长的时光和过去的年代，
组成并填满我们这个世界的事物
一直存在，那么它们肯定
被赋予了不灭的特性，
因此，万物不可能归于乌有。

卢克莱修在论及原子时，认为宇宙之间除原子和空虚之外别无他物存在。物体看起来是静止的，但组成它的原子却在不断地运动着。诗人用远山放牧的羊群来比喻静止的物体，羊群中每只羊的徐徐行动象征着组成物质的原子运动。这是较近似地比喻原子在物体中运动的一幅何等优美的图画啊！诗文中批判了较重的原子落得较快的错误理论；又正确地说明了物体在水中和稀薄空气中以不同速度下落，是由于物体受到不同程度阻碍的缘故。诗人认为原子的形状可能不同；但不同意有巨大尺寸的原子存在。

卢克莱修在论及宇宙的有限和无限问题上，认为宇宙在它的向前道路上没有一个地方是被限制了的。他设想了一个飞矛实验：假使有人走到宇宙的尽头而掷出矛去，诗人巧妙地问道，飞矛将投射到更远的地方去呢？还是被那里的一堵墙挡住射不出去？因此他断言：空间的无限，即使是闪电以无穷时间疾驰也不能把它穿透。空间的无限性，也就是空

间的本性，被诗人用朴素的唯物论和巧妙的比喻说明了。

卢克莱修又论证了物质的无限性。世界是怎样形成的呢？他说：我们的世界不是由于什么心灵的聪明或者像订立契约规定所造成，而是从远古以来物质（原子）遭受冲撞打击，试过了所有各种各样运动和结合之后，终于达到了那些伟大的排列方式。结论是，物质是无限的，运动是永恒的。在论及无限多世界时，诗人认为既有无限多原子在无限的空间里冲撞，就有机会能把世界产生出来。他断言，大地、太阳、月亮、海洋和其他一切，在宇宙中都不是孤单地存在，甚至认为宇宙间的物质在数量上远远超出当时人的计算能力。

在将要结束这首伟大诗篇的时候，诗人再一次以无神论者的声调高唱：

牢记这些，
你就能顿悟，
自然是自由自随，
不受天使们的主宰，
宇宙自己运动，
不烦神仙理睬。
谁能有力量驾驭这不可名数的世界？
哪里有大手能执着这硕大无比的缰和鞭？
谁能够同步同调地旋转天穹？
又谁能煽起天火使大地温暖，
养育出稻、麦、黍、菽无限？
谁能驾风御云使大地顿时昏暗，
驱驰雷霆把自己的庙宇摧残？
霹雳一声又使无辜者遭殃，

逍遥法外的却是坏蛋？

卢克莱修发出一连串的问题，实际上就是对唯心论者的有力批判，对有神论者的尖刻讥讽和无情鞭挞。由于这首不朽的诗篇，马克思赞誉卢克莱修为“朝气蓬勃、叱咤世界的大胆诗人”。

2.3.4 地心说

当时的学者们大都认为各个天体围绕着不动的地球作匀速圆周运动，然而观测中却发现火星存在“留和逆行”，明显不符合匀速圆周运动的原理。一些天文学家企图用圆来“拯救这些现象”。

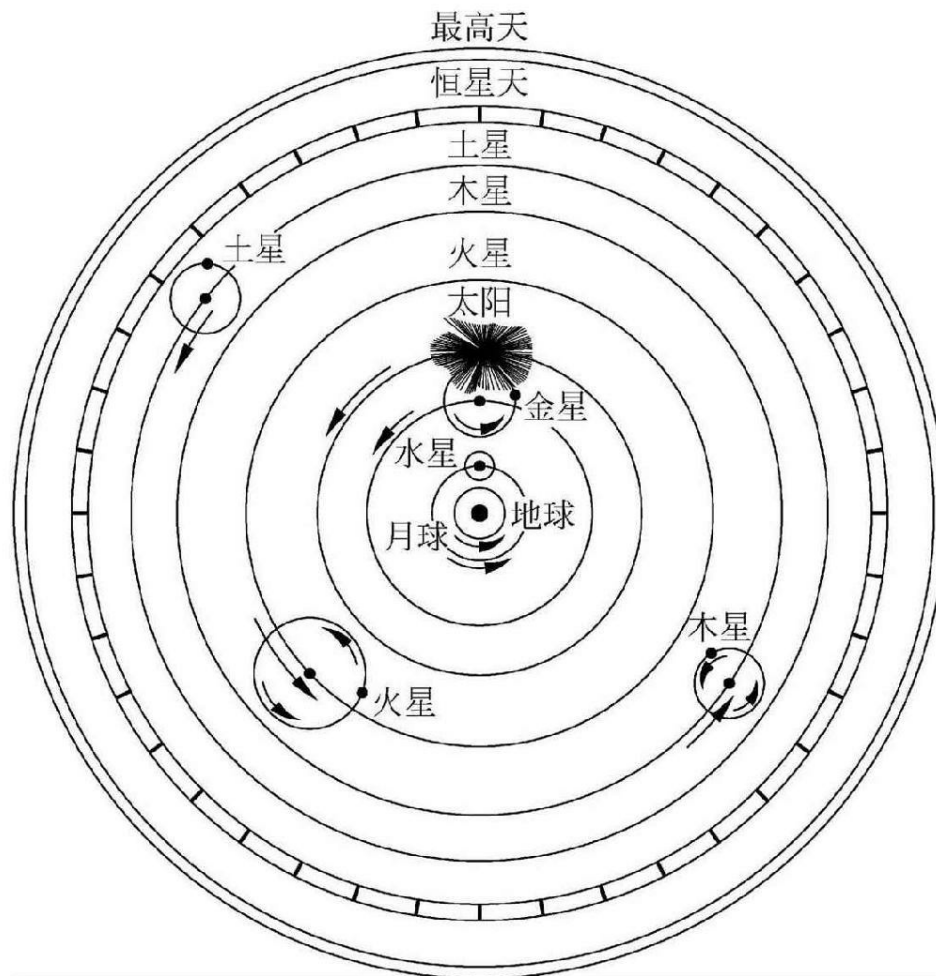
为此，欧多克斯提供了一个解决方案：同心球叠加模型。“欧多克斯发现，用3个球就可以复制出日、月的运动，行星的运动则要用4个球。这样，5大行星加上日、月和恒星天，一共需要27个球。”适当限制同心球的旋转轴、球半径与旋转速度，大部分天体运行情况便可以模拟出来，只是仍有许多较大疏漏。欧多克斯的学说虽然相较于毕达哥拉斯学派主张的地球在运动的学说是后退了一步，但同心圆模式却引导着希帕克斯与托勒密向着更为完整的地心说体系进发。



克罗狄斯·托勒密

公元2世纪，克罗狄斯·托勒密在希帕克斯的学说与研究成果基础

上，结合自己的观测结果，创建了偏心圆与滚圆的运动体系，用以克服同心圆模型的缺陷，并写出古代西方最详尽、最完整的天文学巨作——《至大论》。此后1500年内宇宙学取得的最为瑰丽的成就，莫过于托勒密的《至大论》。托勒密将本轮-均轮与偏心圆体系引入，创造了均衡点，将传统的“地心说”转化为“偏心说”。站在均衡点上的人，可以发现行星在作匀速圆周运动，而事实上它们相对于地球的速度是变化的。托勒密的《至大论》的1~2卷是描述地心体系的基本构造，3~13卷则是在地心体系内讨论诸多天文现象，包括太阳、月球、行星的运动，以及日食、月食的计算方法，恒星和岁差现象。这部天文学的百科全书对地心体系从理论到实际观测都进行了详尽的数学分析，有着极强的接受能力，“能够较好地容纳望远镜出现之前不断出现的新天文观测，所以一直被作为最好的天文学体系，统治西方天文学界一千多年。”



托勒密著有4本重要著作：《至大论》、《地理学》、《天文集》和《光学》。13卷巨著《至大论》直到开普勒的时代，都是天文学家的必读书籍。8卷《地理学》，是他所绘的世界地图的说明书，其中也讨论到天文学原则。他还著有5卷《光学》，其中第1卷讲述眼与光的关系，第2卷说明可见条件、双眼效应，第3卷讲平面镜与曲面镜的反射及太阳中午与早晚的视径大小问题，第5卷试图找出折射定律，并描述了他的实验，讨论了大气折射现象。此外，尚有年代学和占星学方面的著作等。

托勒密的天体模型之所以能够流行千年，是有它的优点和历史原因的。它的主要特点是：

(1) 绕着某一中心作匀速圆周运动，符合当时占主导思想的柏拉图的假设，也适合于亚里士多德的物理学，易于被接受。

(2) 用几种圆周轨道的不同组合预言了行星的运动位置，与实际相差很小，相比以前的体系有所改进，还能解释行星的亮度变化。

(3) 地球不动的说法，对当时人们的生活是令人安慰的假设，也符合基督教信仰。

在当时的历史条件下，托勒密提出的行星体系学说，是具有进步意义的。首先，它肯定了大地是一个悬空着的没有支柱的球体。其次，从恒星天体上区分出行星和日、月是离我们较近的一群天体，这是把太阳系从众星中识别出来的关键性一步。

托勒密本人声称他的体系并不具有物理真实性，而只是一个计算天体位置的数学方案。至于教会利用和维护地心说，那是托勒密死后的事情了。教会之所以维护地心说，只是想歪曲它以证明教义中描绘的天堂、人间、地狱的图像，如果编纂教义时流行着别的什么学说，说不定教会也会加以利用的。所以，托勒密的宇宙学说同宗教本来并没有什么必然的联系。

2.3.5 历法编制、天象观测、记录和天文测量

公元前5世纪，人们已测定了年的长度，并且采用阴阳历。

雅典天文学家默冬创立19年7闰法，称为“默冬章”，确定年长365.2632天，朔望月29.531 92天。

后又有学者确定年长365.2467天，朔望月29.530 85天。

到公元前2世纪，喜巴恰斯提出年长365.2467天，朔望月29.530 59天。

天象观测和记录：古希腊时期有彗星记录，但很不完整。喜巴恰斯于公元前134年在天蝎座发现一颗新星。他通过观测恒星编制星表，包含有1080颗（一说850颗）恒星。他把恒星按星等划分为6等。他还发现岁差，并以太阳轨道偏心来解释。

天文测量：公元前3世纪，有学者测量了日、月大小和距离。方法巧妙，思路很严谨，但测量误差太大。

α 角、日、月、地球大小和距离测定与实际值对比

	测定值	实际值
α 角	3°	10'
日地距离 / 月地距离	(18~20) : 1	约390 : 1
太阳直径 / 地球直径	(6.33~7.33) : 1	约109 : 1

还有学者测定了地球大小。测得塞恩（今阿斯旺）和亚历山大城的距离为地球圆周长的1/50。两地距离为5000希腊里，故地球周长为25万希腊里。1希腊里=158.5米，地球周长合39 625千米。

测量仪器：喜巴恰斯使用过黄道浑仪。

公历的由来：公元前6世纪罗马共和国使用阴阳历，年平均长度365.25天。公元前1世纪中期儒略·凯撒大帝改革历法，使用阳历。公元前46年颁行。历法取年平均长度365.25天。后奥古斯都帝修正了历法，即为现行公历的前身——儒略历。直到公元1582年罗马教皇格里高里十三世颁布改历命令，在历法中作了两项改正，并修订年平均长度为

365.2425天，这就是现行历法——格里历。

2.3.6 小结

古代文明中的宇宙观，无论是“日心说”还是“地心说”，与当今科学界所公认的宇宙学说相比，有着难以弥补的差距。但是，在缺少科学手段的古代，科学家们以高超的智慧和开创精神，在自由、民主、开放、追求真理的氛围下，提出的所有宇宙学理论，都是人类在认识宇宙的历史进程中留下的印迹，有着不可替代的地位。即便是对我们今天而言，古代学者的证实研究方法与创新思维方式都有值得借鉴之处。而任何理论，要经受客观实际的检验，除此之外，别无他法。经不起实践检验和时间考验的，终究是谬误，这是必须时刻牢记的。

参考文献

- [1] 陈美东. 中国古代天文学思想 [M]. 北京: 中国科学技术出版社, 2007.
- [2] 周成华. 先秦文学观止 [M]. 长春: 吉林大学出版社, 2010.
- [3] 杨金适. 荀子史话 [M]. 北京: 人民出版社, 2014.
- [4] 张鸿, 张分田. 王充 [M]. 昆明: 云南出版集团公司, 2009.
- [5] 袁珂. 中国神话传说 [M]. 北京: 世界图书出版公司, 2012.
- [6] 许结. 张衡评传 [M]. 南京: 南京大学出版社, 2011.
- [7] 范晔. 后汉书 [M]. 北京: 中华书局, 2012.
- [8] 雷连城. 涿水历史文化辑萃 [M]. 北京: 中国文史出版社, 2006.
- [9] 百度网. 古印度天文学 [OL]. 2015-03-14.
<http://baike.baidu.com/view/26066.htm>
- [10] 中国科学技术大学天体物理组. 西方宇宙理论评述 [M]. 北京: 科学出版社, 1978.
- [11] [英] 罗素. 西方哲学史 [M]. 何兆武, 译. 北京: 商务印书馆, 1977.
- [12] [古罗马] 卢克莱修. 物性论 [M]. 方书春, 译. 北京: 译林

出版社，2011.

[13] [英] 约翰·D.巴罗. 宇宙之书：从托勒密、爱因斯坦到多重宇宙 [M]. 李剑龙，译. 北京：人民邮电出版社，2013.

[14] 钮卫星. 天文学史 [M]. 上海：上海交通大学出版社，2011.

3 宇宙及天文学挣脱桎梏迈入科学

这个时期，中国的宇宙及天文学由鼎盛滑向衰落，尽管有杰出的科学家不懈努力，也取得了成就，但是由于政治制度腐朽，压制科技文化，总体上停滞不前。而西方的宇宙及天文学却挣脱了中世纪宗教神学的桎梏，进入了科学的轨道，以哥白尼和牛顿为代表，形成近代宇宙及天文学。后来的“西学东渐”，使中西方宇宙及天文学逐步融汇到一起，取代了传统天文学。

3.1 中国的宇宙及天文学从鼎盛滑向衰落

3.1.1 天文历法和测量

自祖冲之的《大明历》后，唐、宋、元、明、清都编制了相应的历法，修正前面历法的偏差，以与天体运行和农时更为吻合。

麟德历：唐高宗麟德二年（公元665年）颁行，由李淳风制订。这部历法废除平朔，采用定朔；废除闰周，由观测和统计确定置闰；改正周日视差对交食影响；还简化了计算。

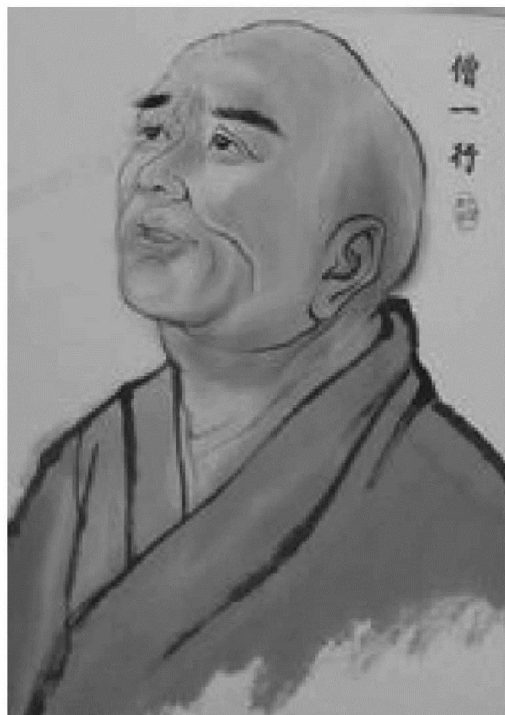
大衍历：唐玄宗开元十五年（公元727年）由一行编订。共有历术7篇、略例3篇、历议9篇，阐述编历内容、日月位置和大行星位置、日月食、夜晚所见恒星等。提出“食差”概念及其计算公式。编排方式成为后代典范。

一行（公元683—727年），本名张遂，是我国古代杰出的天文学家，也是密宗教理的组织者。

开元九年（公元721年），据《麟德历》进行的几次预报日食、月食的时间不准，唐玄宗命一行主持修编新历。从此，一行就开始专门从事天文历法的工作。

开元十一年（公元723年），为了测定星体位置的需要，一行与梁令瓚等人制成了黄道游仪、“水运浑天仪”。当时梁令瓚设计了一个黄道游仪，并已经制成了该仪器的木头模型。在一行的支持和领导下，用铜铸造成此仪器。这台仪器既可以用来测定每天太阳在天空中的位置，也可以用来测定月亮和星宿的位置。

同年，一行和梁令瓚等人在继承张衡“水运浑象”理论的基础上又设计制造了“水运浑天仪”。水运浑天仪上刻有二十八宿，通过注水激轮，每天一周，恰恰与天体周日视运动一致。水运浑天仪的一半在水柜里，一半在水柜的上框，齿轮传动，每逢整时则自然撞钟。整个水运浑天仪既能演示日、月、星辰的视运动，又能自动报时。这是世界上最早的计时器，比外国自鸣钟的出现早了600多年。



僧一行

开元十二年（公元724年），一行根据修改旧历的需要，又组织领导了我国古代第一次天文大地测量，也是一次史无前例、世界罕见的全国天文大地测量工作。

一行主张在实测的基础上修订历法，在经过几年的天文观测及准备工作后，于开元十三年（公元725年）才开始编历。他用两年时间写成历法草稿，并定名为《大衍历》。

《大衍历》以刘焯的《皇极历》为基础，并进一步发展了《皇极历》。《大衍历》共分为7篇，即步中朔术、步发敛术、步日躔术、步月离术、步轨漏术、步交会术、步五星术。《大衍历》发展了前人岁差的概念，创造性地提出了计算食分的方法，发现了不等间距二次内插法公式、新的二次方程式求和公式，并将古代“齐同术”（通分法则）运用于历法计算。

开元十七年（公元729年），《大衍历》颁布实行，并一直沿用达800年之久。经过验证，《大衍历》比当时已有的其他历法，如祖冲之的《大明历》、刘焯的《皇极历》、李淳风的《麟德历》等要精密、准确得多。《大衍历》作为当时世界上较为先进的历法，相继传入日本、印度，在这两国也沿用近百年，极大地影响了这两个国家的历法。

一行在天文方面也做出了重大贡献，他通过长期的天文观测发现了恒星移动的现象，进一步发现和认识了日、月、星辰的运动规律，废弃了沿用长达800多年的二十八宿距度数据，并在历史上第一次提出了月亮比太阳离地球近的科学论点。

从开元十二年（公元724年）起，一行主持了全国范围内的大规模天文大地测量工作。他在全中国选择了12个观测点，并派人实地观测，自己则在长安总体统筹指挥。其中负责在河南进行观测的南宫说等人所测得的数据最科学和有意义。他们选择了经度相同、地势高低相似的4个地方进行设点观测，分别测量了当地的北极星高度，冬至、夏至和春分、秋分四时日影的长度，以及四地间的距离。最后经一行统一计算，得出了南北两地相距351里80步（即现在的129.2千米），北极高度相差1度的结论。虽然这与今天1度对应于111.2千米的测量值相比有较大误差，但这是世界上第一次用科学方法进行的子午线实测，在科学发展史上具有划时代的意义。中国科技史专家李约瑟就曾评价一行组织的子午线长度测量是“科学史上划时代的创举”。

一行在天文历法上所取得的卓越成就在人类文明史上占有重要地位，而且他所重视的实际观测的科学方法，极大地促进了天文学的发展。在他之后，实际观测就成为了历代天文学家从事学术研究时采用的基本方法，引导着学者们译解了一层层的天文奥秘。1000多年后，为纪念这位出色的中国古代天文学家，人们将一颗小行星命名为“一行”。

关于阳历的设想和实践：11世纪末，北宋科学家沈括（公元1031—1095年）在《梦溪笔谈》中提出改阴历为阳历的设想。名之曰《十二气历》。

熙宁五年（公元1072年），沈括负责汴河水建设时，还负责领导司天监，先后罢免了六名旧历官。他不计出身，破格推荐精通天文历算、出身平民的淮南人卫朴进入司天监，主持修订新历的重要工作。沈括和卫朴治



沈括

学态度认真，对旧历官凭借演算凑数的修历方法非常不满，主张从观测天象入手，以实测结果作为修订历法的根据。为此，沈括首先研究并改革了浑仪、浮漏和影表等旧式的天文观测仪器。

浑仪是测量天体方位的仪器。经过历代的发展和演变，到了宋朝，浑仪的结构已经变得十分复杂。三重圆环，相互交错，使用起来很不方便。为此，沈括对浑仪作了比较多的改革。他一方面取消了作用不大的白道环，把仪器简化、分工，再借用数学工具把它们之间的关系联系起来；另一方面又提出改变一些环的位置，使它们不挡住观测视线。沈括的这些改革措施为仪器的发展开辟了新的途径。后来元朝郭守敬于元世祖至元十三年（1276年）创制的新式测天仪器——简仪，就是在这个基础上产生的。

漏壶是古代测定时刻的仪器，由几个盛水的容器装置成阶梯的形式，每一容器下侧都有孔，依次往下一容器滴水漏水。最下面的容器没

有孔，里面装置有刻着时间标度的“箭”，随着滴漏水面升高，“箭”就慢慢浮起，从显露出来的刻度就可以读出时刻。沈括对漏壶也进行了改革。他把曲筒铜漏管改做直颈玉嘴，并且把它的位置移到壶体下部。这样流水更加通畅，壶嘴也坚固耐用多了。

此外，沈括还制造了测日影的圭表，而且改进了测影方法。

他在《浑仪议》、《浮漏议》和《景表议》三篇文章中介绍了自己的研究成果，详细说明了改革仪器的原理，阐发了自己的天文学见解，这在中国天文学史上具有重要的作用。

沈括和卫朴的一系列革新活动遭到守旧势力的攻击和陷害，但在他们的斗争下，卫朴主持修订的奉元历终于在熙宁八年（公元1075年）修成颁行。但是，由于守旧势力的阻挠和破坏，奉元历只施行了18年就被废止了。但是沈括并不因此而灰心，在晚年又进一步提出了用“十二气历”代替原来历法的主张。

中国原来的历法都是阴阳合历，而“十二气历”却是纯粹的阳历。它以十二气作为一年，一年分四季，每季分孟、仲、季三个月，并且按节气定月份，立春那天算一月一日，惊蛰算二月一日，以此类推。大月31天，小月30天，大小月相间，即使有“两小相并”的情况，不过1年只有1次。有“两小相并”的〔解释： $(31+30) \times 6 = 366$ ， $31 \times 5 + 30 \times 7 = 365$ ，故有两小相并的年份。〕，1年共有365天；没有“两小相并”的，1年共366天。这样，每年的天数都很整齐，用不着再设闰月，四季节气都是固定的日期。至于月亮的圆缺，和寒来暑往的季节无关，只要在历书上注明“朔”、“望”就行了。沈括所设计的这个历法是比较科学的，它既符合天体运行的实际，也有利于农业活动的安排。不过，他预见到自己的这一主张必定会遭到守旧派的“怪怒攻骂”、极力阻挠而暂时不能实行，

但是他坚信异时必有用予之说者。

物理学方面的成就：《梦溪笔谈》中所记载这方面的见解和成果，涉及力学、光学、磁学、声学等各个领域。在磁学方面，沈括在《梦溪笔谈》中第一次明确地谈到磁针的偏角问题，沈括还最早发现了地理南北极与地磁场的N、S极并不重合，所以水平放置的小磁针指向跟地理的正南北方向之间有一个很小的偏角；在光学方面，他通过观察实验，对小孔成像、凹面镜成像、凹凸镜的放大和缩小作用等作了通俗生动的论述。他对中国古代传下来的所谓“透光镜”（一种在背面能看到正面图案花纹的铜镜）的透光原因也做了一些比较科学的解释，推动了后来对“透光镜”的研究；此外，沈括还用剪纸人在琴上做过实验，研究声学上的共振现象。

此外，沈括在数学、地理学、医药学和军事方面均有很大成就。

1979年7月1日，中国科学院紫金山天文台为了纪念他，将1964年发现的2027号小行星命名为沈括星。

统天历：南宋宁宗庆元五年（公元1199年）颁行，杨忠辅创制。取年长365.2425日。

授时历：元世祖至元十八年（公元1281年）颁行，郭守敬和王恂等创制，是古历中最精良的历法，采用当时最精确的数据。施行最久，历时364年。采用三次差内插法计日、月运动；用类似球面三角方法（弧矢割圆术）作日、月位置的坐标换算；采用百进制简化运算。

郭守敬（公元1231—1316年），元朝著名的天文学家、数学家、水利专家和仪器制造专家。字若思，汉族，顺德邢台（今河北省邢台市邢台县）人。郭守敬曾担任都水监，负责修治元大都至通州的运河。1276

年，郭守敬修订新历法，经4年时间制订出了《授时历》。

在《授时历》里，有许多革新创造的成绩。第一，废除了过去许多不合理、不必要的计算方法，例如避免用很复杂的分数来表示1个天文数据的尾数部分，改用十进小数等。第二，创立了几种新的算法，例如三差内插内式及合于球面三角法的计算公式等。第三，总结了前人的成果，使用了一些较进步的数据，例如采用南



郭守敬

宋杨忠辅所定的回归年，以一年为365.2425日，与现行公历的平均一年时间长度完全一致。《授时历》是1281年颁行的，而现行公历却是到1852年才由教皇格列高列颁行。《授时历》是我国古代一部很进步的历法。它流传到后世，把许多先进的科学成就传授给后人。这件工作，称得起是郭守敬的一个大功。

郭守敬于公元1276年创制了一种测量天体位置的仪器。该仪器在结构和使用上都比浑仪简单，而且除北极星附近以外，整个天空一览无余，故称简仪。简仪的这种结构，同现代称为“天图式望远镜”的构造基本上是一致的。在欧洲，像这种结构的测天仪器，要到18世纪以后才开始从英国流传开来。

郭守敬简仪的刻度分划空前精细。以往的仪器一般只能读到1度的 $\frac{1}{4}$ ，而简仪却可读到1度的 $\frac{1}{36}$ ，精密度一下提高了很多。这架仪器一直到清初还保存着，可惜后来被当废铜销毁了。现在只留下一架明朝正统年间（1436—1449）的仿制品，保存在南京紫金山天文台。

郭守敬用这架简仪作了许多精密的观测，其中的两项观测对于新历的编算有重大的意义。

一项是黄道和赤道交角的测定。赤道是指天球的赤道。地球悬空在天球之内，设想地球赤道面向周围伸展出去，和天球边缘相割，割成一个大圆圈，这个圆圈就是天球赤道。黄道就是地球绕太阳作公转的轨道平面延伸出去，和天球相交所得的大圆。天球上黄道和赤道的交角，就是地球赤道面和地球公转轨道面的交角，这是一个天文学基本常数。这个数值从汉朝以来一直认定是 24° ，一千多年来始终没有人怀疑过。实际上这个交角年年在不断缩减，只是每年缩减的数值很小，只有半秒，短时间不觉得。可是变化虽小，积累了1000多年也会显示出影响。黄、赤道交角数值的精确与否，对其他计算结果的准确与否有很大影响。因此，郭守敬首先对这个沿用了千年的数据进行检查。果然，经他实际测定，当时的黄、赤道交角只有 $23^{\circ}90'$ 。这个是用古代角度制算出的数目。古代把整个圆周分成 365° ， 1° 分作100分，用这样的记法来记这个角度就是 $23^{\circ}90'$ 。换成现代通用的 360° 制，那就是 $23^{\circ}33'23''.3$ 。根据现代天文学理论推算，当时的这个交角实际应该是 $23^{\circ}31'58''.0$ 。郭守敬测量的角度实际还有 $1'25''.3$ 的误差。不过这样的观测，在郭守敬当年的时代来讲，已是难能可贵的了。

另一项观测就是二十八宿距度的测定。我国古代在测量二十八宿各个星座的距离时，常在各宿中指定某处星为标志，这个星称为“距星”。因为要用距星作标志，所以距星本身的位置一定要定得很精确。从这一宿距星到下一宿距星之间的相距度数叫“距度”，它可以决定这两个距星之间的相对位置。二十八宿的距度，从汉朝到北宋，一共进行过5次测定。它们的精确度是逐次提高的。最后1次在宋徽宗崇宁年间（公元1102—1106年）进行的观测中，这28个距度数值的误差平均为 $0^{\circ}.15$ ，也

就是9′。到郭守敬时，经他测定的数据，误差数值的平均只有4′.5，比崇宁年间的那一次降低了一半。这也是一个很难得的成绩。

简仪的主要装置是由两个互相垂直的大圆环组成，其中的一个环面平行于地球赤道面，叫做“赤道环”。另一个是直立在赤道环中心的双环，能绕一根金属轴转动，叫做“赤经双环”。双环中间夹着一根装有十字丝装置的窥管，相当于单镜筒望远镜，能绕赤经双环的中心转动。观测时，将窥管对准某颗待测星，然后在赤道



简仪

环和赤经双环的刻度盘上直接读出这颗星星的位置值。有两个支架托着正南北方向的金属轴，支撑着整个观测装置，使这个装置保持着北高南低的形状。这是我国首先发明的赤道装置，要比欧洲人使用的赤道装置早500年左右。

在编订新历时，郭守敬提供了不少精确的数据，这是新历得以成功的一个重要原因。

在改历过程中，郭守敬创造了近20种仪器和工具。我们再介绍一件郭守敬独创的仪器，来看看他的技术成就。

这件仪器是一个中空的铜制半球面，形状像一口仰天放着的锅，名叫“仰仪”。半球的口上刻着东西南北的方向，半球口上用一纵、一横的两根竿子架着一块小板，板上开一个小孔，孔的位置正好在半球面的球心上。太阳光通过小孔，在球面上投下一个圆形的像，映照在所刻的线格网上，立刻就可以读出太阳在天球上的位置。这样，人们就可以不用

眼睛逼视光度极强的太阳本身，就可以看明白太阳的位置，这是很巧妙的。更妙的是，在发生日食时，仰仪面上的日像也相应地发生亏缺。这样，从仰仪上就可以直接观测日食的方向，亏缺部分的多少，以及发生各种食像的时刻等。虽然伊斯兰天文家在古时候就已经利用日光通过小孔成像的现象观测日食，但他们只是利用一块有洞的板子来观测日面的亏缺，帮助测定各种食像的时刻罢了，还没有像仰仪这样可以直接读出数据的仪器。

郭守敬等人同一位尼泊尔的建筑师合作，在大都（北京）兴建了一座新的天文台，台上安置有郭守敬所创制的那些天文仪器。这是当时世界上设备最完善的天文台之一。



望远镜

由于郭守敬的建议，朝廷派了14位天文家，到当时国内26个地点（大都不算在内）进行了几项重要的天文观测。在其中的6个地点，特别测定了夏至日的表影长度和昼、夜的时间长度。这些观测结果，都为编制全国适用的历法提供了科学的数据。这一次天文观测的规模之大，在世界天文学史上也是少见的。

公元1279年（己卯年），郭守敬奉旨进行“四海测验”，在南海的测量点就在中国黄岩岛。

后来，郭守敬升为太史令。在以后的几年间，他又继续进行天文观测，并且陆续地把自己制造天文仪器、观测天象的经验和结果等极宝贵的知识编写成书。他写的天文学著作共有百余卷之多。然而封建帝王元

世祖虽然支持了改历的工作，却并不愿让真正的科学知识流传到民间去，他把郭守敬的天文著作统统锁在了深宫秘府之中。封建的专制，使那些宝贵的科学遗产几乎全都被埋没了，使我国的宇宙及天文学逐步由鼎盛滑向衰落，这是多么令人痛惜的事。

1981年，为纪念郭守敬诞辰750周年，国际天文学联合会以他的名字为月球上的一座环形山命名，并将2012号小行星命名为“郭守敬”。

河南登封至今仍保留有郭守敬所建的观星台。

《崇祯历书》和《西洋新历书》与时宪历：

明崇祯二年（公元1629年）徐光启组织历局，聘请西洋传教士罗雅各、汤若望等参加编译，崇祯七年编成。《崇祯历书》采用第谷宇宙体系和几何学的计算系统，引入地理经纬度，应用球面三角学，采用通行的角度和时间度量单位制。开始与西方宇宙及天文学进行融合。

清顺治元年（公元1644年）汤若望将此书更名为《西洋新法历书》，他也同时被任命为钦天监监正，编成的时宪历完全采用定朔和定气。

徐光启（公元1562—1633年），字子先，号玄扈，教名Paul（保罗），汉族，松江府上海县人，中国明末的数学家、科学家、农学家、政治家、军事家。官至礼部尚书、文渊阁大学士，是中西文化交流的先驱之一。徐光启在天文学上的成就主要是主持历法的修订和《崇祯历书》的编译。

编制历法，在中国古代是关系到“授民以时”的大事，为历代王朝所重视。由于中国古代数学历来以实际计算见长，重视和历法编制之间的

关系，因此中国古代历法的准确程度是比较高的。但是到了明末，却明显地呈现出落后的状态。一方面是由于西欧的天文学此时有了飞速进步，另一方面则是明王朝长期执行不准私习天文，严禁民间研制历法政策的结果。封建专制严重阻碍了科学技术的发展。明代施行的《大统历》，实际上就是元代《授时历》的继续，日久天长，已严重不准。据《明史·历志》记载，自成化（公元1481年）年间开始，陆续有人建议修改历法，但建议者不是被治罪便是以“古法未可



徐光启

轻变”、“祖制不可改”为由遭到拒绝。万历三十八年（公元1610年）11月发生的日食，司天监再次预报错误。徐光启以西法推算最为精密，礼部奏请开设历局。以徐光启督修历法，改历工作终于走上正轨。但后来清朝入主中原，改历工作在明代实际并未完成。

《崇祯历书》的编译，自崇祯四年（公元1631年）起直至十一年（公元1638年），才编译完成。全书46种、137卷，是分五次进呈的。前三次乃是徐光启亲自进呈（23种、75卷），后两次都是徐光启死后由李天经进呈的。其中第四次是徐光启亲手订正（13种、30卷），第五次则是徐氏“手订及半”最后由李天经完成的（10种、32卷）。

《崇祯历书》采用的是第谷体系。这个体系认为地球仍是太阳系的中心，日、月和诸恒星均作绕地运动，而五星则作绕日运动。这比传教士刚刚到达中国时由利玛窦所介绍的托勒密体系稍有进步，但对于当时

西方已经出现的更为科学的哥白尼体系，传教士则未予介绍。《崇祯历书》仍然用本轮、均轮等一套相互关联的圆运动来描述和计算日、月、五星的疾、迟、顺、逆、留、合等现象。对当时西方已有的更为先进的行星三大定律（开普勒三定律），传教士也未予介绍。尽管如此，按西法推算的日、月食精确程度已较中国传统的《大统历》更高。此外《崇祯历书》还引入了大地为球形的思想，大地经纬度的计算及球面三角法，区别了太阳近（远）地点和冬（夏）至点的不同，采用了蒙气差修正数值。他为中国天文界引进了星等的概念；根据第谷星表和中国传统星表，提供了第一个全天性星图，成为清代星表的基础。

徐光启在数学方面也取得杰出成就，概括地说，有三个方面，即：

- （1）论述了中国数学在明代落后的原因；
- （2）论述了数学应用的广泛性；
- （3）与意大利传教士利玛窦一起翻译并出版了《几何原本》。

然而，明朝时《几何原本》并没有得到重视，致使徐光启逝世后《几何原本》迟迟不能继续翻译完成，以至于被埋没。后来明朝灭亡，清统治者对此书并不关注。康熙皇帝虽然重视西学，但是很可惜《几何原本》这样重要的著作还是没用到，没能完成徐光启的遗愿。再次看出封建专制制度的落后对科学文化的阻碍作用。

3.1.2 宇宙观念和思想

唐宋学者们的宇宙思想也很有见解。

沈括认为“天地之变，寒暑风雨，水旱螟蝗，率皆有法”，并指

出，“阳顺阴逆之理，皆有所从来，得之自然，非意之所配也”。就是说，自然界事物的变化都是有规律的，而且这些规律是客观存在的，是不以人们的意志为转移的。他还认为事物的变化规律有正常变化和异常变化之分，不能拘泥于固定不变的规则。正是这些比较正确的思想观点，促使他取得了那个时代在科学技术方面达到的高度成就。沈括提出已知的知识是有限的，人的认识是无限的观点，对科学的发展产生了较大的影响。

“天”是什么？按照神秘哲学的看法，“天”是神灵、上帝，宇宙间的一切是上帝神灵创造的。

柳宗元（公元773—819年），唐代文学家，学者，从元气一元论的自然哲学观出发，彻底否定了上帝神灵说。他在《天对》中的回答是：“黑晰眇，往来屯屯，庞昧革化，惟元气存，而何为焉！”柳宗元认为，日月昼夜，交替运行，永不停息，宇宙从蒙昧混浊的状态变化发展产生万物，只是因为“元气”存在的缘故，哪里是由谁造成的呢？在他眼里，“天”是自营自成的自然界，是宇宙，不是虚无缥缈的神；“天”是物的天，是客观存在的天，没有意识和感知；“天”并不是谁经营创造的，而是由元气的无限积聚而形成的；元气是一切自然现象发生之源，一切自然现象统一于元气。这一思想从根本上否认了造物主的存在，坚持了世界的物质性，表明了鲜明的无神论立场。

南宋思想家朱熹则从哲学体系范畴论述了他的宇宙思想。他认为：天地初间，只是阴阳之气。这一筒气运行，磨来磨去，磨得急了，便拶许多渣滓，里面无处出，便结成个地在中央。气之清者便为天，为日月，为星辰，只在外周环运转。地便只在中央不动，不是在下。



柳宗元



朱熹

朱熹（公元1130—1200年）是南宋著名的理学家、思想家、哲学家、教育家、诗人，闽学派的代表人物，世人尊称“朱子”。他曾先后在

庐山建立“白鹿洞书院”，在武夷山修建“武夷精舍”，在岳麓山修建“岳麓书院”进行讲学。而武夷山被列入世界文化及自然双重遗产，朱子的文化讲学是个重要因素。他继承儒家，兼采释、道各家思想，形成了一个庞大的哲学体系，其中就包含着他的宇宙观念和思想。这一体系的核心范畴是“理”，或称“道”、“太极”。朱熹所谓的“理”，有几方面互相联系的含义：①“理”是先于自然现象和社会现象的形而上者。他认为“理”比“气”更根本，逻辑上“理”先于“气”；同时，“气”有变化的能动性，“理”不能离开“气”。②“理”是事物的规律。③“理”是伦理道德的基本准则。④“理”在人身上就是人性。朱熹又称“理”为“太极”，是天地万物之“理”的总体，即总万物的那个“理”。“太极”既包括万物之“理”，万物便可分别体现整个“太极”。这便是人人有一“太极”，物物有一“太极”。每一个人和物都以抽象的“理”作为它存在的根据，每一个人和物都具有完整的“理”，即“理一”。“气”是朱熹体系中仅次于“理”的第二个范畴。它是形而下者，是有情、有状、有迹的；它具有凝聚、造作等特性。它是铸成万物的质料。天下万物都是“理”和质料相统一的产物。朱熹认为“理”和“气”的关系有主有次。“理”生“气”并寓于“气”中，“理”为主，为先，是第一性的；“气”为客，为后，属第二性。

朱熹不信鬼神，不相信世间万物是由鬼神主宰的，更不相信冥冥之中是有定数的。他认为逢事在人为，没有人不可以做到的事情。

动静观：朱熹主张“理”依“气”而生物，并从“气”展开了一分为二、动静不息的生物运动，这便是一“气”分做二“气”，动的是阳，静的是阴，又分做五行（金、木、水、火、土），散为万物。一分为二是从“气”分化为物过程中的重要运动形态。朱熹认为由对立统一，而使事物变化无穷。他探讨了事物的成因，把运动和静止看成是一个无限连续的过程。时空的无限性又说明了动静的无限性，动静又是不可分的。这

表现了朱熹思想的辩证法观点。朱熹还认为动静不但相对待、相排斥，并且相互统一。他还论述了运动的相对稳定和显著变动这两种形态，他称之为“变”与“化”。他认为渐化中渗透着顿变，顿变中渗透着渐化。渐化积累，达到顿变。

格物致知论：朱熹用“致知在格物”的命题，探讨认识领域中的理论问题。“格物致知”是他认识论的核心，把道德看作天道的体现。即通过道德修养，追求“至诚”的境界，以感应天地，达到“天人合一”。在认识来源问题上，朱熹既讲人生而有知的先验论，也不否认见闻之知。他强调穷理离不得格物，即格物才能穷其理。朱熹探讨了“知”“行”关系。他认为“知”先“行”后，“行”重“知”轻。从知识来源上说，“知”在先；从社会效果上看，“行”为重。而且“知”“行”互发，“知之愈明，则行之愈笃；行之愈笃，则知之益明”。

从此，朱熹开始建立自己的一套客观唯心主义思想——理学。朱熹认为在超现实、超社会之上存在一种标准，它是人们一切行为的标准，即“天理”。只有去发现（格物穷理）和遵循天理，而破坏和谐的是“人欲”。因此，他提出“存天理，灭人欲”。这就是朱熹客观唯心主义思想的核心。淳熙三年（1176年），朱熹与当时著名学者陆九渊相会于江西上饶鹅湖寺，交流思想。但陆属主观唯心论，他认为人们心中先天存在着善良，主张“发明本心”，即要求人们自己在心中去发现美好事物，达到自我完善。这与朱的客观唯心说的主张不同。因此，二人辩论争持，以至互相嘲讽，不欢而散。这就是中国思想史上有名的“鹅湖会”。从此有了“理学”与“心学”两大派别。

朱熹的《朱子家训》：“勿以善小而不为，勿以恶小而为之。人有恶，则掩之；人有善，则扬之。处世无私仇，治家无私法。勿损人而利己，勿妒贤而嫉能。勿称忿而报横逆，勿非礼而害物命。见不义之财勿

取，遇合理之事则从。诗书不可不读，礼义不可不知。子孙不可不教，童仆不可不恤。斯文不可不敬，患难不可不扶。守我之分者，礼也；听我之命者，天也。人能如是，天必相之。此乃日用常行之道，若衣服之于身体，饮食之于口腹，不可一日无也，可不慎哉！”（节选自《紫阳朱氏宗谱》）

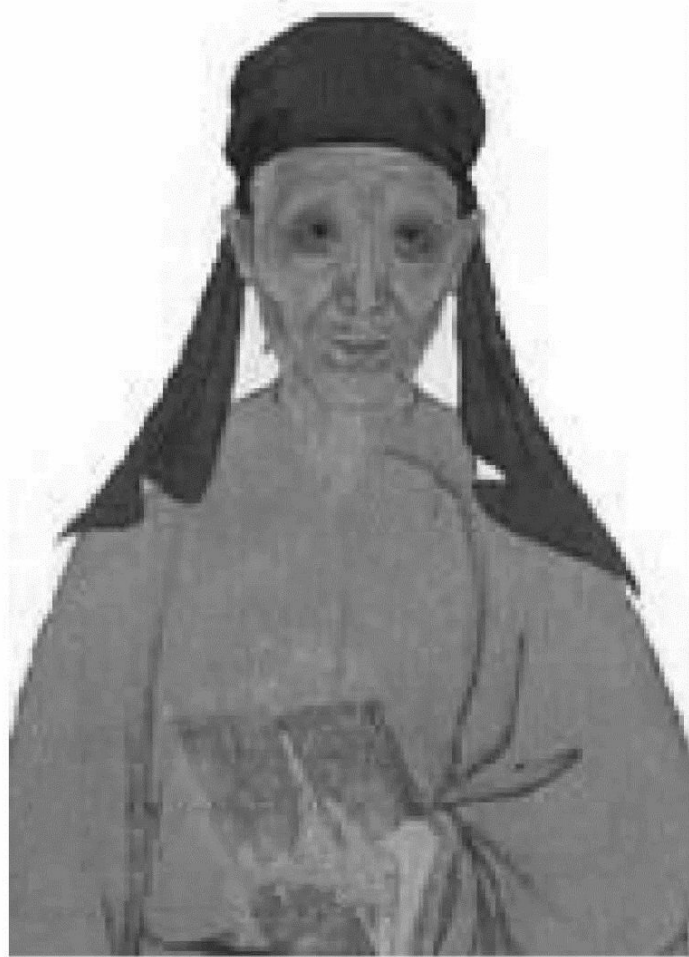
他的学说，不仅成为中国的国学，而且从14世纪开始，就已经相继流传于日本、朝鲜等东南亚诸国。朝鲜李氏王朝非常推崇朱子学说，而日本从德川幕府时代起，就以朱熹的学说为官学了。他的学说对我国封建社会的历史发展影响很大，以致后来的康熙皇帝都称朱熹为：“集大成而绪千百年绝传之学，开愚蒙而立亿万世一定之归。”

朱熹虽然认为“天理为义，人欲为利”，但并不一概反对功利。他的基本态度与孔子一样，是重义轻利，以公利至上。希望人们要“见利思义”，甚至“舍生取义”。他还发挥了孟子的思想，把“明人伦”称作“明义理以修其身”，继而做到“修身、齐家、治国、平天下”。

以上介绍了古代大学者朱熹的宇宙思想和理学，这是历史的一段过程。对于我们今天，应该是取其精华，去其糟粕，这才是正确的。



王夫之



黃宗羲



顾炎武

王夫之（公元1619—1692年），汉族，字而农，号姜斋，生于今衡阳市雁峰区，明末清初伟大的思想家、文学家、史学家兼美学家。王夫之学识极其渊博，经学、子学、史学、文学、政法、伦理等各门学术，造诣无不精深，天文、历数、医理、兵法乃至卜筮、星象，亦旁涉兼通，且留心当时传入的“西学”。

他认为，整个宇宙除了“气”，更无他物。他还指出“气”只有聚散、往来，而没有增减、生灭。所谓有无、虚实等，都只是“气”的聚散、往来、屈伸的运动形态。按当时科学发展水平，他举例论证了“气”的永恒不灭性，认为这种永恒无限的“气”乃是一种实体；并提出“太虚，一实

者也”、“充满两间，皆一实之府”等命题，力图对物质世界最根本的属性进行更高的哲学抽象。他把“诚”训为“实有”，以真实无妄的“实有”来概括物质世界的最一般属性。他还认为，客观世界万事万物的本质和现象都是客观实在的，“从其用而知其体之有”、“日观化而渐得其原”，可以通过认识各种物质现象而概括出它们的共同本质。从而否定了唯心主义空无本体的虚构。

在“理”与“气”的关系问题上，王夫之坚持“理依于气”的气本论，驳斥了以“理”为本的观点。他强调“气”是阴阳变化的实体，“理”乃是变化过程所呈现出的规律性。“理”是气之“理”，“理”外没有虚托孤立的“理”。从而批判了周敦颐、朱熹等所坚持的“气”外求“理”的唯心主义理论。王夫之结合对“统心、性、天于理”的客观唯心主义体系的批判，强调指出：“盖言心言性，言天言理，俱必在气上说，若无气处，则俱无也。”明确地坚持了唯物主义的氣本论。提出“理即气之理，而后天为理之义始成”，有力地批判了宋明理学的“理在气先”、“理在事先”，即精神先于物质存在的唯心论，否认了离开物质运动而独立存在的客体精神——理。他发展了“理也顺而不妄”的观点，说明了“理”不仅在“气”中，而且是“气”的运动变化，有它的“必然”——规律性。

王夫之坚持“无其器则无其道”、“尽器则道在其中”的唯物主义道器观，系统地驳斥了割裂、颠倒“道”“器”关系的唯心主义思想。他给传统“道”与“器”范畴以新的解释，认为“形而上”的“道”与“形而下”的“器”所标志的一般（共同本质、普遍规律）和个别（具体事物及其特殊规律），两者是“统此一物”的两个方面，是不能分离的。他提出“天下惟器而已矣”的命题，肯定宇宙间一切事物都是具体的存在，任何具体事物都具有特殊本质，又具有同类事物的共同本质，“道者器之道”，一般只能在个别中存在，只能通过个别而存在，“终无有虚悬孤致

之道”。犹如没有车马便没有御道，没有牢醢、璧币、钟磬、管弦便没有礼乐之道一样。他明确指出，在“器”之外、“器”之先安置一个“无形之上”的精神本体，乃是一种谬说。他通过论证“道”对于“器”的依存性，得出了“据器而道存，离器而道毁”的结论，驳斥了“理在事先”、“道本器末”的观点。

王夫之对“道”“器”关系作了新的发展。他说：“据器而道存，离器而道毁”。所谓“器”，就是指客观存在的各种具体物质，所谓“道”是具体事物的规律；没有事物，运动的规律就是不存在的，所以“道不离器”。还认为，“无其器则无其道”，即没有事物就没有事物的规律，只能说规律是事物的规律，而绝不能说事物是规律的事物。总之，当有某种事物的时候才会有关于它的原则、道理和规律。同时还认为，随着“器日尽，而道愈明”，意思是说，随着事物向前发展，它所表现的规律也就愈明显了。他的“道不离器”的观点，坚持了物质第一性，精神第二性的唯物主义观点。

在“知”“行”关系问题上，他力图全面清算“离行以为知”的认识路线，注意总结不同学派长期争鸣的思想成果，在理论上强调“行”在认识过程中的主导地位，得出了“行可兼知，而知不可兼行”的重要结论。他以“知”源于“行”、力“行”而后有真知为根据，论证“行”是“知”的基础和动力、“行”包括“知”、统率“知”。同时，他仍强调“知行相资以为用”。王夫之进一步提出“知之尽，则实践之”的命题，认为“可竭者天也，竭之者人也。人有可竭之成能，故天之所死，犹将生之；天之所愚，犹将哲之；天之所无，犹将有之；天之所乱，犹将治之”。人可以在改造自然、社会 and 自我的实践中，发挥重大作用。这种富于进取精神的朴素实践观，是王夫之认识论的精华，为其唯物主义体系奠定了基础。

王夫之以唯物主义自然观去观察历史，提出“理”、“势”统一的历史

观。他把历史发展的客观过程和必然趋势，叫做“势”，把历史发展的规律性叫做“理”，提出了“于势之必然处见理”的观点，即人们必须从“势之必然处”认识历史发展的必然规律。他还进一步提出，历史既然有“理”和“势”，治天下就必须“循理”、“乘势”，按照客观规律办事。因此，他强调，历史发展不能凭主观意志，而必须遵守历史发展的客观规律。同时，他还重视人的能动作用。他认为，从一种客观可能性变为社会现实，必须通过人的有目的的活动。

在发展观方面，王夫之综合以往丰富的认识成果，并对自己所面对的复杂的社会矛盾运动进行哲学概括，对中国古代辩证法的理论发展做出了重要贡献。

王夫之与宋明以来流行的主静说相对立，而坚持主动论，提出“物动而已”、“动以入动，不息不滞”、“天地之气，恒生于动而不生于静”，把自然界看作永恒运动化生着的物质过程。他否定了周敦颐、朱熹所宣扬的“太极动静而生阴阳”的观点，指出：“动而生阳，动之动也，静而生阴，动之静也，废然无动而静，阴恶从生哉”。说明运动是物质世界所固有的，否定从气以外寻找事物运动原因的外因论。他针对“静为躁君”、“静非对动”的动静观，明确肯定“静由动得”而“动静皆动”。但他并不否认静止的意义和作用，以为相对的静止是万物得以形成的必要条件。阳变阴合的运动过程本身包含着动静两态：绝对的动，相对的静。这样，否定了主静说，又批判了割裂动静的各种形而上学的运动观，更深一层地阐述了动静两者的辩证联系。

王夫之强调“天地之化日新”，把荣枯代谢、推移吐纳看作是宇宙的根本法则。他认为任何生命体都经历着胚胎、流荡、灌注、衰减、散灭诸阶段，前三者是生长过程，后二者是衰亡过程，而就在“衰减”、“散灭”过程中已经孕育“推故而别致其新”的契机，旧事物的死亡准备了新

事物诞生的条件，“由致新而言之，则死亦生之大造矣”。这种变化发展观，有一定的理论深度，并富于革新精神。

王夫之把事物运动变化的原因，明确地归结为事物内部的矛盾性，认为“万殊之生，因乎二气”。他在“一物两体”学说的基础上开展了他的矛盾观，提出“乾坤并建”、“阴阳不孤行于天地之间”，肯定矛盾的普遍性。对于矛盾着的对立面之间的关系，他进一步分析指出，任何矛盾都是相反相成的，一方面“必相反而相为仇”，这是排斥关系；另一方面“相反而固会其通”，这是同一关系。这两重关系，不可分割，“合二以一者，就分一为二之所固有”。但他更强调“由两而见一”，认为矛盾双方绝非截然分开，而是“反者有不反者存”。按他的分析，“阴阳者，恒通而未必相薄，薄者其不常矣”。矛盾双方互相逼迫、激烈搏斗的状态是“反常”的，而互相联合、贯通，保持同一性状态才是“正常”的。在他看来矛盾是相互转化的，有时会发生突变，但在更多的情况下，转化是在不断往复、消长中保持某种动态平衡而实现的。

王夫之的辩证发展观，尤其是他的矛盾学说，具有重要的理论价值，但他过分强调矛盾的同一性，则是有局限性的。

王夫之在物质运动问题上，认识到物质运动的绝对性，批判形而上学不变论。他说：“天下之变万千，而要归于两端。”意思是说，世界变化无穷无尽，究其原因，是由于气中存在着两端。“两端”就是事物存在的两个方面，比如阴和阳、刚和柔、动和静、聚和散等。任何一个事物都包含着这“两端”。他认为静与动的关系是辩证的，他说：“静者静动，非不动也”、“方动即静，方静施动，静即含动，动不舍静”。这就是说，动是绝对的，静是相对的，如江河之水，表面看来似乎古今一样，其实今水已非古水。他认为“天地万物，恒生于动而不生于静”，他还阐述了“道日新”、“质日代”的发展变化观点。他说：“天地生物，其

化不息”，是说事物是永远发展变化的，不可能“废然而止”。他还修正了“日月之形，万古不变”的观点。

黄宗羲（公元1610—1695年）是明末清初经学家、史学家、思想家、地理学家、天文历算学家、教育家，汉族，浙江绍兴府余姚县人，字太冲，一字德冰，号南雷。他反对“理在气先”的理论，认为“理”并不是客观存在的物质实体，而是“气”的运动规律，认为“气质人心是浑然流行之体，公共之物也”，具有唯物论的特色。“盈天地皆心也”的观点又有唯心论的倾向，是一种混合的宇宙观念。黄宗羲的政治思想是近代民主思想，在民权理论上已超越了欧洲的卢梭。

顾炎武（公元1613—1682年），汉族，著名思想家、史学家、语言学家，与黄宗羲、王夫之并称为明末清初三大思想家。现江苏省昆山市人，本名绛，字忠清。他所提出的“天下兴亡，匹夫有责”这一口号，意义和影响深远，成为激励中华民族奋进的精神力量。在顾炎武的一生中，也确实是以“天下为己任”而奔波于大江南北。即使他在病中，还在呼吁“天生豪杰，必有所任。……今日者，拯斯人于涂炭，为万世开太平，此吾辈之任也”。他提倡经世致用，反对空谈，注意广求证据，提出“君子为学，以明道也，以救世也。徒以诗文而已，所谓雕虫篆刻，亦何益哉？”“能于政事诸端切实发挥其利弊，可谓内圣外王体用兼备之学”，“礼义廉耻，是谓四维”，“不廉则无所不取，不耻则无所不为”。这些至理名言我们应时时谨记。

3.1.3 世界上的首次航天实践

古时的火箭是将火药装在纸筒里，然后点燃，发射出去。起初只是用于过年、过节放烟火时使用，是我国首先发明的。到了13世纪，人们

把火箭用作战争武器，后来后传入了欧洲。

第一个想到利用火箭飞天的人是我们中国人——明朝的万户。



传说中第一个使用火箭的人——万户

美国火箭学家赫伯特·S.基姆在1945年出版的《火箭和喷气发动机》一书中提到，“约当14世纪之末，有一位中国的官吏，官职为万户，但其姓名没有明文记载，因此后人把他叫做万户了。他在一把座椅的背后，装上47枚当时能买到的最大火箭。他把自己捆绑在椅子的前边，两只手各拿一个大风筝，然后叫他的仆人同时点燃47枚大火箭，其目的是想借火箭向上推进的力量，加上风筝上升的力量飞向上方。他的目标是月亮！”

关于明朝人万户的故事，我国古代的文献资料虽然记述不多，但当时的场景在600多年后的今天看来，仍是那样的惊心动魄、令人赞叹。据记载，万户是明朝初期人，原来是一个木匠。由于他喜欢钻研技巧，尤其是对技术发明方面特别痴迷，所以从军后改进过不少当时军队里的刀枪车船。

万户的本领是在明王朝同瓦剌的战事中被发现的。同样对兵器制造很有研究的明朝大将军班背认为，正是因为万户对武器的改良才使得战争取得根本胜利，所以奏请朝廷让万户到兵器局供职。当时，中国四大发明之一的火药已经在军事上初露锋芒，所以万户的前途本该是一片光明的。

但可惜的是，和万户相交甚好的班背将军性情耿直，从不趋炎附势，因得罪右中郎李广太等奸臣而被革职，并幽禁在拒马河上游的深山鬼谷中。为了从深山里营救出好友班背将军，聪明的万户决定造一只“飞鸟”。但由于其他因素，将军被政敌杀害，救人的计划落空。失去了知己的万户这个时候厌恶了官场和人世间的生活，于是他开始谋划着逃离是非官场和人间，决定到月球上去生活。

在那个人类对自然界认识受到很大局限的特殊时代，木匠出身的万户甚至做出了一份很详尽的科学理论计算报告，他认为按照当时的火箭技术，再加之风筝原理的帮助，他一定能在一个时间段内飞到月亮上去。在这个理想主义者的思维世界里，月亮上是没有人险恶的……

为了实现自己的意愿，同时也是为了实现将军班背的遗愿，万户开始潜心研究将军遗留下来的《火箭书》，并用自己的知识给予完善。他造出了各种各样的火箭，然后画出飞鸟的图形，众匠人按图制造出了飞鸟……

在一个月明如盘的夜晚，万户带着人来到一座高山上。他们将一只巨大的“飞鸟”放在山头上，“鸟头”正对着明月……万户拿起风筝坐在“鸟背”上的驾驶座位——椅子上。他先点燃“鸟尾”引线，一瞬间，火箭尾部喷火，“飞鸟”离开山头向前冲去。接着万户的两只脚下也喷出火焰，“飞鸟”随即又冲向半空。

但不幸的是，后来人们在远处的山脚下发现了万户的尸体和“飞鸟”的残骸……这个故事后来被记载为“万户飞天”。万户虽然失败了，但他借助火箭推力升空的创想是世界上第一个，因此他被世界公认为“真正的航天始祖”，为了纪念这位世界航天始祖，20世纪70年代，国际天文学联合会将月球背面一座环形山命名为“Wan Hoo”（“万户”）。

3.1.4 小结

这一时期中国的宇宙及天文学在以一行、沈括和郭守敬为代表的科学、技术工作者的奋力实践下达到鼎盛，取得了巨大成就；宇宙思想及观念也在以王充、朱熹、王夫之和顾炎武为代表的学者勤奋开拓下颇有建树，但没有形成理论体系。然而，在封建专制、腐败制度的束缚下，它们却逐步滑向衰落，以致落后于西方。一些觉醒的人，在来华传教士的帮助下，开始向西方学习。但受到的阻力很大，收效甚微。

3.2 西方的宇宙及天文学挣脱桎梏，走入科学轨道

3.2.1 中世纪的宗教神学扼杀宇宙及天文学的发展

公元4、5世纪间，欧洲的封建社会开始建立。之后，又一直延续了大约1000多年，史称“中世纪”。中世纪欧洲社会的特点是政教合一，教廷、教宗掌握绝对权力，不但统治社会百姓，还要控制人的思想。要让人们信仰、膜拜他们宣扬的那一套。宗教神学思想成为占统治地位的正统思想。宗教神学在认识论上主张精神第一、上帝（天主）万能，因而强烈反对科学和一切不同的思想。在此期间，西罗马主教奥古斯丁提出哲学应服务于神学、理性应服从信仰的观点。他对基督教教义的布道和信仰成为了中世纪早期的文化基础。他也借用柏拉图的哲学，企图来阐明基督教教义。他的言论很有影响，一直到公元1200年还广为引用，比如灵魂为“圣光”所照耀，灵魂领悟存在于上帝的万物“永恒意志”等，甚至自然科学家也深受影响。12世纪有位研究光学的英国物理学家曾探究虹的成因，并在作出虹是阳光在云层中发生折射的科学解释的同时，还在创立和奥古斯丁“永恒意志”、“圣光”之类谬说相一致的、形而上学的光学理论。由此可以看出，奥古斯丁的神学影响之广远。5世纪以后，欧洲人连希腊哲学家的名字都不知道，诵读的都是神学家的经书，论证的都是圣经中上帝创造世界的奇迹，信奉的都是主教奥古斯丁和教父的神学说教，把“正因为荒谬，我才相信”奉为信条。有的人甚至把地圆说也列为异端，无知地讥讽说：“如果地是圆的，就会有部分人头在下，脚朝上。在那里雨雪由下而升，田野、海洋、城镇、山岳不用绳子悬挂也不落下，简直比巴比伦的空中花园还神奇。”

中世纪的欧洲产生了有代表性的经院哲学，又称“僧侣主义”，阿奎那（1225—1274）是其代表人物。他有句名言：“哲学是神学的婢女。”在他的神学里，上帝至上、精神第一，万能的上帝创造了宇宙一切。即真理不靠感性认识，也不靠理性认识，而是靠对上帝的信仰，靠上帝给的“天启”。也就是说，人的正确思想不是从实践中来的，而是完全靠上帝的恩赐得来的。这正是纯粹的先验论。正因为如此，阿奎那的形而上学神学体系成了罗马教廷的官方哲学。

12世纪中叶，古希腊学者亚里士多德著作的阿拉伯译本从西班牙传入法国后，又有其他古希腊学者的著作译本，相继被转译成拉丁文。阿奎那和他的老师还要求把一些著作直接从希腊文译过来。这些经院哲学家们似乎急于要学习希腊哲学。但事实上，阿奎那之辈偷窃并歪曲了亚里士多德书中的一部分，企图以亚里士多德的话来论证天主教义，使哲学匍匐在神台之下，充当神学的奴仆。他们还把托勒密体系和天主教教义相结合，随心所欲地引证托勒密的话来为其神学服务。

从公元前直到公元后的一段时间中，由于天文观测水平的限制，亚里士多德—托勒密的地心说体系尚能符合那时的天文观测情况，在当时的生产实践上起到一定作用。但进入中世纪后，该体系为宗教神学所窃取和篡改之后，地心说的宇宙理论就彻底变成了为神学服务的僧侣主义，充当了中世纪政教合一的统治集团的帮凶。对于前后两个时期中亚里士多德—托勒密体系所起的作用是根本不能等量齐观的。可以说，到了中世纪，从古代开始的关于宇宙结构的争论，就以披着亚里士多德—托勒密体系外衣的宗教神学获得全面统治，而暂时告一段落。

在宗教神学的严重影响下，14世纪左右盛行的宇宙结构是怎样的呢？大地的下面是地狱，是囚禁有罪之人的地方。炼狱耸立于大海之上的高峰上，人的灵魂必须爬过七层炼狱以求赎罪；然后才是天上的乐

园，即天堂。宗教的“圣地”耶路撒冷则居于大地的中央。这就是中世纪黑暗时期，宗教用以愚弄人们的荒谬的宇宙图景。

宗教神学和经院哲学对宇宙及天文学的停滞起了主要作用，占星术转向迷信、滑入歧途，也起到了不好的作用。其实应该说，占星术对古代天文学的发展也有一定的促进作用。为了进行星占，人们注意观测天象，并留下了丰富的天象记载。但后来，西方的占星术逐渐发展到对个人进行星占。例如，根据一个人诞生时日、月、五星在黄道十二宫中的位置，推算“算命天宫图”，以占卜个人一生的命运。占星术牵强附会地把天象与人事联系在一起，有些国王还把占星学家视为高参，往往请他们根据星象占卜来确定对重大政治、军事事件的决策，这就毫无科学可谈了。

3.2.2 复兴的前奏

进步、科学的思想和正确的观点是不会屈服的。公元13世纪，英国有个革新派教徒罗杰·培根（约1214—1293），是个具有唯物主义倾向的哲学家和自然科学家，著名的唯名论者，实验科学的前驱。他具有广博的知识，素有“奇异的博士”之称。他怀疑推理演绎法，坚持以实验来验证的可靠性，对光性质和虹的研究颇有独到之处，他还绘制了眼镜的制作原理，阐述了反射、折射、球面光差的原理和机械推动船只和车辆的原理。他利用镜子和透镜在炼金术、天文学与光学中进行实验，是第一位讲述如何制造弹药的欧洲人。他对经院哲学进行批判，反对经院式、教义式的盲目信仰，认为只有实验科学才能“造福人类”，对宇宙理论和科学进展起了促进作用。

培根把自然界作为哲学研究的主要对象，并强调知识最根本的来源

是经验。他说，“认识”有3种方法：权威、判断和实验。权威必须通过理智来判断，而判断又必须通过实验才能证实是真理，所以人类认识的道路“是从感官知识到理性”，“没有经验就不能充分认识任何事物”。他严厉地斥责对权威的盲目崇拜，以及经院哲学家的因循守旧、不学无术和空洞烦琐的论证，认为这是认识真理的四大障碍。在宗教神学占绝对统治地位的中世纪，这些思想体现出他的勇敢战斗精神。他公然怀疑旧约全书上的话，声称圣经上的话要以科学来证明后才能相信。他也因此触犯了教廷，晚年一直被禁锢在监狱里。不过罗杰·培根和神学的对抗只是一场大风暴的前奏。

培根的这些观点并不彻底。他曾宣称科学研究得越充分，就越能论证神学，认为“神圣的启示”和“内在的启发”也属经验之列，并且是“认识”的更好的途径。这是历史条件和他个人的宗教生活在他身上打下的思想烙印。

欧洲的中世纪是个“黑暗的时代”。基督教教会建立了一套严格的等级制度，把上帝当作绝对的权威。文学、艺术、哲学一切都得遵照基督教的经典——《圣经》的教义，谁都不可违背，否则宗教法庭就要对他制裁，甚至处以死刑。在教会的管制下，中世纪的科学技术进展非常缓慢。黑死病在欧洲的蔓延，也加剧了人们心中的恐慌，使得人们开始怀疑宗教神学的绝对权威。

中世纪后期，资本主义萌芽在生产力的发展等多种条件的促生下，于欧洲的意大利首先出现。这是商品经济发展到一定阶段的产物，此时自由的贸易开始呼唤人的自由，陈腐的欧洲需要一场新的提倡个人自由的思想运动。

资本主义萌芽的出现为这场思想运动的兴起提供了可能。城市经济

的繁荣，使事业成功财富巨大的富商、作坊主和银行家等更加相信个人的价值和力量，更加充满创新进取、冒险求胜的精神。多才多艺、高雅博学之士受到人们的普遍尊重。这为文艺复兴的发生提供了深厚的物质基础和适宜的社会环境以及人才。

在古希腊和古罗马，文化艺术的成就很高，人们可以自由地发表各种学术思想，这和黑暗的中世纪是个鲜明的对比。14世纪末，由于信仰伊斯兰教的奥斯曼帝国不断入侵东罗马（拜占庭），东罗马人带着大批的古希腊和罗马的艺术珍品和文学、历史、哲学等书籍，纷纷逃往西欧避难。一些东罗马的学者在意大利的佛罗伦萨办了一所叫“希腊学院”的学校，讲授希腊辉煌的历史文明和文化等。这种辉煌的成绩与资本主义萌芽产生后的成果，显示出其优越性远远高于黑暗的中世纪，使得人们追求更自由的社会环境和思想精神氛围。

于是，许多欧洲的学者要求恢复古希腊和罗马的文化和艺术。这种要求就像春风，慢慢吹遍了整个欧洲。文艺复兴的先驱但丁（1265—1321）创作了大量文学作品，其代表作为《神曲》。他的作品以含蓄的手法批评和揭露中世纪宗教统治的腐败和愚蠢，首次以意大利方言而不是作为中世纪欧洲正式文学语言的拉丁文进行创作，被认为是资产阶级叩响近代社会大门的文艺复兴运动代表第一人。他认为古希腊、罗马时代是人性最完善的时代，中世纪将人性压制是违背自然的。

当时的意大利处于城邦林立的状态，各城市都是一个独立或半独立的城邦国家，14世纪后各城市逐渐从共和制走向独裁。独裁者耽于享乐，信奉新柏拉图主义，希望摆脱宗教禁欲主义的束缚，大力保护艺术家对世俗生活的描绘。与此同时圣方济各会的宗教激进主义力图摒弃正统宗教的经院哲学，转而歌颂自然的美和人的精神价值。同时，罗马教廷也在走向腐败，历届教皇的享乐规模比世俗独裁者还要厉害。他们也

在保护艺术家，允许艺术偏离正统的宗教教条。哲学、科学都在逐渐地朝着比较宽松的氛围发展，一场宗教改革正在逐渐酝酿起来。

改革者积极提倡人文主义精神，其核心是以人为中心而不是以神为中心，肯定人的价值和尊严；主张人生的目的是追求现实生活中的幸福，倡导个性解放，反对愚昧迷信的神学思想，认为人是现实生活的创造者和主人。

在诸多因素的合力之下，欧洲近代三大思想解放运动（文艺复兴、宗教改革与启蒙运动）由此兴起。也给欧洲的宇宙及天文学带来了复兴的曙光。

3.2.3 欧洲及周边阿拉伯地区的宇宙、天文学的复兴

由于地域相邻、往来密切的关系，欧洲与阿拉伯地区的宇宙、天文学已经互相交汇、融合在一起。公元7世纪初，阿拉伯的圣人穆罕默德（570—632）创立伊斯兰教，统一了阿拉伯各国。在宇宙及天文学方面，先后形成多个学派，使其得到了很大发展。

（1）巴格达学派

公元661年，倭马亚王朝（白衣大食）时期，建立大马士革天文台。

公元750年，阿拔斯王朝（黑衣大食）时期，公元829年建立巴格达天文台。

学者巴塔尼（858—929），修改《天文学大成》，撰写《萨比历数

书》，共57章。他发现太阳近地点进动，对欧洲天文学产生了深远影响。

学者苏菲（903—986）编撰了《恒星星座》。

（2）开罗学派

公元909—1171年北非的法提玛王朝（绿衣大食）。

学者伊本·尤努斯（？—1009），撰写《哈基姆历数书》，共81章。

学者伊本·海桑（965—1038），研究球面像差、透镜放大率和大气折射，对欧洲科学发展产生了深远影响。

（3）西阿拉伯学派

建于西班牙的后倭马亚王朝（白衣大食）。

学者查尔卡利元（？—1100）于1080年编成《托莱多天文表》，在欧洲长期行用，提出水星按椭圆轨道运行，否定本轮、均轮学说。

学者伊本·图法（1110—1185）和比特鲁吉·伊什比利提出行星运动几何模型。

学者伊本·鲁什德（1126—1198）指出托勒密体系为非物理的现实。

（4）蒙古统治时期

公元1258年建伊尔汗国。

学者纳西尔丁·图西（1201—1274）建议在伊朗西北部建马拉盖天文台，仪器在当时首屈一指。公元1271年完成《伊尔汗历数书》。

公元1370年建帖木儿帝国。

乌鲁伯格（1394—1449）继位后建撒马尔罕天文台，有口径40米的象限仪。公元1447年编成《乌鲁伯格历数书》，其中包含了一个有1018颗恒星的星表，精度极高。

学者伊本·萨蒂尔（约1305—1375）修正托勒密体系，引入本轮套本轮的体系。

（5）西班牙卡斯蒂利亚国王阿方索十世（1221—1284）支持天文发展，刊布《阿方索天文表》。

（6）法国天文学家霍利伍德出版《天球论》（公元1220年）。

（7）维也纳大学教授波伊尔巴赫出版《行星新论》（公元1474年）。

（8）学者雷乔蒙塔努斯写成《天文学大成概要》（公元1463年）。他自建天文台，编制星历表并出版（公元1474年），还认识到地球运动。

（9）奥里斯姆（约1320—1382）巴黎冲力学派代表人物，论证了地球自转，受原始冲力推动而无限期持续运动。

（10）古萨的尼古拉主教（1401—1464）主张万物运动，宇宙无限。

3.2.4 哥白尼日心体系的创立和发展

16世纪中叶，哥白尼日心体系创立，标志着近代天文学诞生。为此做出重大贡献的人物包括：哥白尼、布鲁诺、第谷、开普勒、伽利略、牛顿等著名的科学家。

1543年，近代科学的开山之作——哥白尼的《天体运行论》——终于出版。尽管哥白尼本人生性怯懦，直到晚年才有胆量将作品交付刊印，但这并不影响他成为科学史上的科学巨人身份。

哥白尼的《天体运行论》并非是对希腊科学体系的完全否定。无论是从他的个人履历，还是它所采用的科学理论，从某种角度上都可以说，“哥白尼只是重新倡导希腊人的传统知识”。

哥白尼将宇宙的中心置于太阳附近，认为所有的天体都围绕着太阳公转，取消了偏心匀速点，使得整个体系更为简洁、美观，同时也解决了行星的“留和逆行”问题。他还将地球置于地月系统的中心，认为月球环绕地球运动。哥白尼通过精确计算认为，日地距离和天穹高度之比远小于地球半径与日地距离之比，因此得出日地距离相较于天穹高度微不足道的结论，很好地解决了恒星周年视差的问题。在此书中，哥白尼还指出，天穹上呈现的任何运动形式，并非是天穹本身的运动，而是地球自转与公转共同作用的结果。

这些理论看似极为简洁，哥白尼却进行了相当详尽复杂的数学分析与逻辑论证，显示了他作为“新柏拉图主义者”对于完美的追求，这也为近代科学开创了数理天文学传统。实际，哥白尼对他的宇宙体系的描述，还不到《天体运行论》的一半篇幅。在剩下的篇幅中，哥白尼运用那令人眼花缭乱的数学公式，试图向人们展示他的宇宙系统。相较于托

勒密体系，“它在智识上更加优雅——更加赏心悦目——而且更加经济。”“就这样在十几个世纪之后，托勒密终于在自己的游戏中失败了。”

关于《天体运行论》：

（1）体例

用拉丁文写成，共6卷。书名为《论天体旋转的六卷集》，后人简称为《天体运行论》。

（2）内容简介

第一卷为宇宙总结构图像，鸟瞰式地介绍了宇宙的结构。在论证的开始，哥白尼列举了许多观测资料来证明地球是圆形的。接着他指出了地球呈圆形的理由。他说：“所有的物体都倾向于将自己凝聚成为这种球状，正如同一滴水或一滴其他的流体一样，总是极力将自己形成一个独立的整体。”“物体呈球状的原因在于它的重量，即在于物体的微粒或者说原子的一种自然倾向，要把自己凝聚成一个整体，并收缩成球状。”

关于原子，他还写了这样一段：“所谓原子，是最细微的、不能再分割的微粒，它们重叠地或是成倍地相聚在一起，但由于它们看不见，所以并不立即形成看得见的物体；可是它们的数量可以增加到这种程度，足够累积到可以看见的大小。”这一段话是针对唯心主义者的论调而说的，他们借口“原子无法看见”而抹杀原子的存在。

第二卷为应用球面三角方法解释天体在天球上的视运动。介绍了有关的数学原理，其中平面三角和球面三角的演算方法都是哥白尼首创

的。这里陈述了三角形的规则，即从三角形的已知某些边和角去推算其他边和角的规则。这包括了三边是直线的平面三角形和三边是球面上圆弧做成的球面三角形。

第三卷为太阳视运动的计算方法。

第四卷为关于月球的视运动。

第五、六卷是关于行星的视运动。

（3）《天体运行论》出版的意义

①动摇了宗教统治的理论支柱。

②自然科学向教会发布的“独立宣言”。

③天文学跨入近代科学大门。为其他科学发展打开通道。

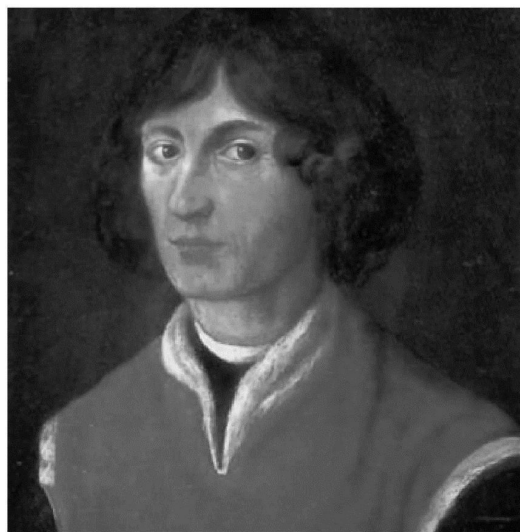
哥白尼革命的意义不仅仅局限于天文学，也不仅仅是人类认识水平上的飞跃，更重要的是引起了西方人价值观念上的转变。

（4）《天体运行论》的缺陷

认为宇宙有限，太阳位于宇宙中心，保留“恒星天”概念。恪守天体只能作匀速圆周运动的观念，保留本轮和均轮体系。

尼古拉·哥白尼于1473年2月19日出生在波兰的托伦。10岁丧父，由舅父抚养。18岁进入克拉科夫大学学习。在听取了布鲁捷夫斯基的天文课后，开始考虑地球的运转问题。他在后来写成的《天体运行论》的序言里说过，前人有权虚构圆轮来解释星空的现象，他也有权尝试发现一种比圆轮更为妥当的方法，来解释天体的运行。

哥白尼观测天体的目的和过去的学者相反。他不是强迫宇宙现象服从“地球中心”学说。哥白尼有一句名言：“现象引导天文学家。”他正是要让宇宙现象来解答他所提出的问题，要让观测到的现象证实一个新创立的学说——“太阳中心”学说。这种目标明确的观测，终于促成了天文学的彻底变革。



哥白尼

哥白尼的观测工作在克拉科夫大学时就有了良好的开端。他曾利用著名的占星家玛尔卿·布利查赠送给学校的“捕星器”和“三弧仪”观测过月食，研究过浩瀚无边的星空。

22岁时，他获得弗龙堡大教堂神父团职位。

23岁时去意大利博洛尼亚大学学习教会法，并师从著名天文学家诺法拉，研究天文学。在这里，哥白尼结识了当时知名的天文学家多米尼克·玛利亚，同他一起研究月球理论，开始用实际观测来揭露托勒密学说和客观现象之间的矛盾。他发现托勒密对月球运行的解释，一定会得出一个荒谬的结论：月亮的体积时而膨胀时而收缩，满月是膨胀的结果，新月是收缩的结果。1497年3月9日，哥白尼和玛利亚一起进行了一次著名的观测。那天晚上，夜色清朗，繁星闪烁，一弯新月高挂天空。他们站在圣约瑟夫教堂的塔楼上，观测“金牛座”的亮星“毕宿五”，看它怎样被逐渐移近的蛾眉月所淹没。当“毕宿五”和月亮相接而还有一些缝隙的时候，“毕宿五”很快就隐没起来了。他们精确地测定了“毕宿五”隐没的时间，计算出确凿不移的数据，证明那些缝隙都是月亮亏食的部分，“毕宿五”是被月亮本身的阴影所遮掩的，月球的体积并没有缩小。

就这样，哥白尼把托勒密的地心说打开了一个缺口。

28岁返回波兰，加入神父团。同年夏天，在意大利帕多瓦大学学医，期间又获费拉拉大学教会法规博士学位。

33岁回到波兰，在埃尔蒙兰其舅父处任职，继续天文研究。他在钻研古代典籍的时候，曾抄下这样一些大胆的见解：

“天空、太阳、月亮、星星以及天上所有的东西都站着不动，除了地球以外，宇宙间没有什么东西在动。地球以巨大的速度绕轴旋转，这就引起一种感觉，仿佛地球静止不动，而天空却在转动。

“大部分学者都认为地球静止不动，但是费罗窝斯和毕达哥拉斯却叫它围绕一堆火旋转。

“在行星的中心站着巨大而威严的太阳，它不但是时间的主宰，不但是地球的主宰，而且是群星和天空的主宰。”

这些古代学者的卓越见解，在当时被认为是“离经叛道”的，但是对哥白尼来说，却好比是夜航中的灯塔，照亮了他前进的方向。

40岁（1513年3月）时，他在弗龙堡大教堂的平台上安置了天文仪器，自此开展天文观测，长达二十年，最终完成了《天体运行论》。约在1515年前，哥白尼为阐述自己关于天体运动学说的基本思想，撰写了篇题为《浅说》的论文，他认为天体运动必须满足以下七点：不存在一个所有天体轨道或天体的共同的中心；地球只是引力中心和月球轨道的中心，并不是宇宙的中心；所有天体都绕太阳运转，宇宙的中心在太阳附近；地球到太阳的距离同天穹高度之比是微不足道的；在天空中看到的任何运动，都是地球运动引起的；在空中看到的太阳运动的一切现

象，都不是它本身运动产生的，而是地球运动引起的，地球同时进行着几种运动；人们看到的行星向前和向后运动，是由于地球运动引起的。地球的运动足以解释人们在空中见到的各种现象了。

此外，哥白尼还描述了太阳、月球、三颗外行星（土星、木星和火星）和两颗内行星（金星、水星）的视运动。书中，哥白尼批判了托勒密的理论。他较科学地阐明了天体运行的现象，推翻了长期以来居于统治地位的地心说，并从根本上否定了基督教关于上帝创造一切的谬论，从而实现了天文学中的根本变革。他正确地论述了地球绕其轴心运转、月亮绕地球运转、地球和其他所有行星都绕太阳运转的事实。但是他也和前人一样严重低估了太阳系的规模。他认为星体运行的轨道是一系列的同心圆，这当然是错误的。他的学说里的数学运算很复杂，但也很不准确。即使这样，他的书还是立即引起了极大的关注，驱使其他天文学家对行星运动作更为准确的观察，其中最著名的是丹麦的天文学家第谷·布拉赫，开普勒就是根据他积累的观察资料，最终推导出了星体运行的正确规律。

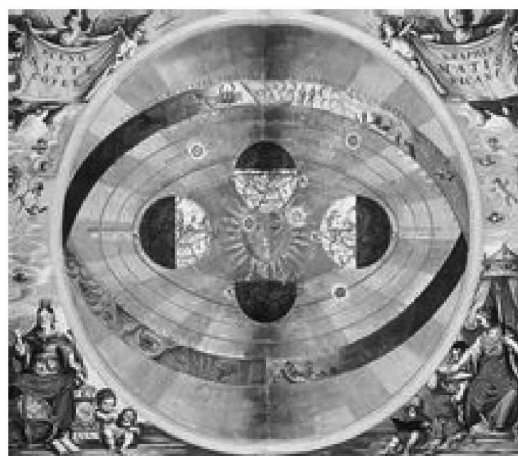
由于朋友们不断催促，哥白尼把他的“太阳中心学说”写出了一个提纲，取了一个朴素的名字，叫《试论天体运行的假设》，抄送给他的几个心腹朋友。它宣布：“所有的天体都围绕着太阳运转，太阳附近就是宇宙中心的所在。地球也和别的行星一样绕着圆周运转。它一昼夜绕地轴自转一周，一年绕太阳公转一周……”。

哥白尼所宣布的是一个巨大的学说体系的轮廓，它在参加聚会的朋友中间引起了许多争论，哥白尼对许多疑问都做了解答。在结束辩论的时候，他引用了古罗马大诗人西塞罗的话：“没有什么东西赶得上宇宙的完整，赶得上德行的纯洁。”他用这句话表明了一具信念，那就是：宇宙是完整的、对称的、和谐的，是具有可以理解的规律和秩序的。

后来哥白尼将他未来的著作取名为《运行》。在他看来，运动才是生命的真谛——运动存在于万物之中，上达天空，下至深海。没有什么东西是静止的，一切东西都在生长、变化、消失，千秋万代继续不停。

《运行》这一著作，就是要揭示大自然这一最本质的秘密。哥白尼的这一观点，肯定了客观世界的存在和它的规律性，闪耀着朴素的唯物主义哲学的光辉。

哥白尼对地球的形状，曾多次作过间接观测。早在1500年11月6日，他就在罗马近郊的一个高岗上观测过月食，研究地球投射在月球表面的弧状阴影，从而证实了亚里士多德关于地球呈球状的论断。在定居弗隆堡时，他曾多次站在波罗的海岸边观察帆船。有一次，哥白尼请求一艘帆船在桅顶绑上一个闪光的物体，他站在



哥白尼绘制的宇宙图

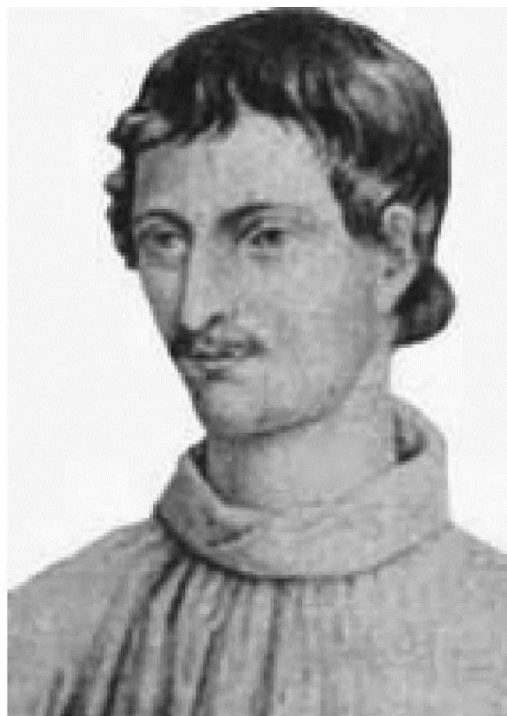
岸边看着这艘帆船慢慢驶离。他描写这次观察的情况说：“随着帆船的远去，那个闪光的物体逐渐降落，最后完全隐没，好像太阳下山一样。”这次观察使他得出一个结论：“就连海面也是圆形的。”

而教廷一直把哥白尼的学说和科学研究视为异端邪说，教区的主教和宗教裁判官长期迫害、打压哥白尼，他们监视、查禁、威胁，用尽各种卑劣手段，妄图让哥白尼屈服。但是，在支持者帮助下，哥白尼克服了重重阻力，他的科学巨著还是出版了。

1543年5月24日，哥白尼用手摸着刚刚送来的《天体运行论》，在病榻上逝世。

3.2.5 为确立哥白尼日心说继续斗争

继哥白尼之后，伽利略、布鲁诺、第谷、开普勒等享誉盛名的科学家们，沿着哥白尼所开创的研究道路，将人类对宇宙的认识，提升到了前所未有的高度，从多个角度证明了哥白尼体系的优越性与正确性。“假如在他死后150年间，没有出现一系列的天才，将他的工作完成，取得他所没有得到的决定性证据，天文学便不会发生伟大的进展，而他的体系也不会流传到今天。”



乔尔丹诺·布鲁诺

必须着重提出的是不屈的斗士——乔尔丹诺·布鲁诺（1548—

1600）。罗马教廷对于刚刊印的《天体运行论》一书采用的是不闻不问的办法。由于哥白尼的著作是用拉丁文写的，而且数学计算很多，只有懂数学的人才能看懂，所以在市民阶层中影响不大。因此，罗马教廷在70多年间没有对哥白尼的著作明令取缔。

布鲁诺在读过《天体运行论》后，发现奥塞安德尔在书中的序言是一篇伪作。他愤慨地指出：“这是一个其蠢如驴、不学无术、狂妄自大的角色，给哥白尼的著作附加一张废纸。”布鲁诺曾先后到过欧洲十几座著名的城市，宣传哥白尼的理论，借以打击神学的世界观。他去讲学的地方有信奉新教的国家，也有信奉天主教的国家，不过即使被人驱逐，他也毫不在乎。但是，当布鲁诺发表著作介绍哥白尼学说的时候，

教廷就采用了严厉镇压的手段。这是因为，布鲁诺写的是明白晓畅、人人都懂的意大利文，文笔优美而犀利，讽刺了那些拼命维护“地球中心学说”的神学家，并揭示了哥白尼学说的全部唯物主义的意义。布鲁诺推崇哥白尼为最伟大的天文学家。他在《哥白尼的光辉》一诗中写道：“你的思想没有被黑暗世纪的卑怯所玷污，你的呼声没有被愚妄之徒的叫嚣所淹没，伟大的哥白尼啊，你的丰碑似的著作在青春初显的年代震撼了我们的心灵。”

他在文章中写道：“我们对哥白尼感激不尽，因为他把我们从居于统治地位的庸俗哲学中解放出来，……只有那种坚定不移地站在反宗教的潮流中的人，才能充分评价并颂扬他的精神。……他给自己找到坚实的立场，并毫不含糊地宣称：承认地球对于宇宙的运动，终究是一件不可避免的事，因为这比认定无数天体（其中许多都比地球更为光辉而庞大）把地球当作中心的说法，要合情合理得多。”

布鲁诺联系古代唯物主义者学说，到处宣传哥白尼的革命理论，而且在宇宙的无限性和运动的永恒性方面发展了哥白尼的学说。他在哥白尼学说的基础上，形成了自己的崭新的宇宙论。他提出并论证了宇宙无限和世界众多的思想。他认为整个宇宙是无限大的，根本就不存在固定的中心，也不存在界限。而地球只是绕太阳运转的一颗行星，太阳也只是宇宙中无数恒星中的一颗。在无限的宇宙中，有无数的“世界”在产生和消亡，但作为无限的宇宙本身，是永恒存在的。布鲁诺不仅抛弃了地球中心说，而且也跨过了哥白尼的太阳中心说而大大前进了一步。他还提出天地同质说，认为物质是一切自然现象共同的统一基础。

布鲁诺明确指出自然界的万事万物都处在普遍联系和不断运动变化之中。这一变化是统一的物质实体包含的各种形式不断转化的过程，事物经过相互转化，形成对立面的统一。布鲁诺还论述了“极大”与“极

小”的对立统一。他指出“宇宙里面，体积与点无别，中心与周边无别，有限者与无限者无别，最大者与最小者无别”。他把对立统一原则看作是认识自然、发现真理的诀窍，将这一学说提到方法论的高度。他得出的结论是：“谁想要认识自然的最大秘密，那就请他去研究和观察矛盾和对立面的最大和最小吧。深奥的法术就在于能够先找出结合点，再引出对立面。”布鲁诺把这种辩证思想推广应用于社会和日常生活。他说：“不可能有这样的国家、这样的城市、这样的世代、这样的家庭，其成员竟会有相同的脾胃，而没有互相对立、互相矛盾的性格。”他指出意大利既是“一切罪恶”的“渊源”，又是“地球的头脑和右手”以及一切美德的“教导者、培育者和母亲”。布鲁诺继承和发展了古代辩证法成为文艺复兴时期伟大的辩证理论家。他提出若干重要辩证的原理并做了详细论证，为反对中世纪经院哲学中行而上学的观点起到了重要作用。

由于布鲁诺在欧洲广泛宣传他的新宇宙观、反对经院哲学，进一步引起了罗马宗教裁判所的恐惧和仇恨。1592年，罗马教徒将他诱骗回国，并逮捕了他。刽子手们用尽种种刑罚仍无法令布鲁诺屈服。他说：“高加索的冰川，也不会冷却我心头的火焰，即使像塞尔维特那样被烧死也不反悔。”他还说：“为真理而斗争是人生最大的乐趣。”经过8年的残酷折磨后，布鲁诺被处以火刑。1600年2月17日凌晨，罗马塔楼上的悲壮钟声划破夜空，传进千家万户。这是施行火刑的信号。通往鲜花广场的街道上站满了群众。布鲁诺被绑在广场中央的火刑柱上，他向围观的人们庄严地宣布：“黑暗即将过去，黎明即将来临，真理终将战胜邪恶！”最后，他高呼“火，不能征服我，未来的世界会了解我，会知道我的价值。”刽子手用木塞堵上了他的嘴，然后点燃了烈火，布鲁诺在熊熊烈火中英勇就义。

还应提到的是观测天文大师——丹麦天文学家第谷·布拉赫（1546

—1601）和天空的立法者——德国天文学家开普勒（1571—1630）。

第谷当初很崇拜哥白尼，他曾经打发人访问弗龙堡，并取回一幅哥白尼的自画像和他生前用过的一架“捕星器”。第谷看到这个仪器时大为惊诧，哥白尼竟是用这么简陋的仪器来考察“天体的奥妙”。他毕恭毕敬地把哥白尼的遗像供在上位，还在遗像下题词：

“力大无比的巨人能够搬过一座山来加到另一座山上，可是雷的劈击却能把巨人制服——比起所有这些巨人，哥白尼一个人不知要坚强多少、伟大多少、幸福多少。他把整个地球连同所有的山岳举起来迎向群星，雷的劈击却不能把他制服。”

第谷在天文观测方面很有成就，1572年11月11日黄昏发现新星，第二年出版《论新星》。1746年，30岁的第谷在汶岛建天文台——天堡和星堡。他还设计制造了多种天文仪器，测量天体位置的精度达到2'。他还坚持多年测定行星视位置、探讨宇宙体系理论。51岁时，第谷离开汶岛，并于53岁到达布拉格，在城郊建天文台进行观测。

1600年2月，开普勒拜访第谷，并在第谷手下研究天文学。次年，第谷临终前把行星运动资料交于开普勒。他们两人的合作是科学史上实测工作与理论工作紧密配合，取得辉煌成功的范例。

开普勒衷心信奉哥白尼的学说。他也曾指出奥塞安德尔在《天体运行论》这本书中伪撰序言，宣扬不可知论的卑劣手段。他还利用自己发明的望远镜，把有关宇宙结构的科学向前推进了一步。

1571年12月27日，开普勒生于德国南部魏尔市。17岁时，靠奖学金进入蒂宾根大学，师从麦斯特林学习天文学。20岁获硕士学位，留校学习神学。23岁受蒂宾根大学评议会推荐，至奥地利格拉茨地方基督教福

音学校任教，同时编算星占历书。

1. 探索宇宙奥秘

开普勒认为6个行星所在的天球球面正好外接于和内切于5种正多面体，发现

土星——正立方体——木星——正四面体——火星——正十二面体——地球——正二十面体——金星——正八面体——水星

这种设计得到的各个球的半径比率与各个行星轨道大小的已知值相当吻合，但没有跳出旧的圆形轨道的框框，若干年后，随着观测研究的深入。开普勒就摒弃了它。

1596年，出版《宇宙的神秘》，受到第谷欣赏。

2. 行星运动三定律的发现

1600年2月，开普勒赴布拉格；1601年，接受第谷临终托付。

开普勒首先研究了地球在轨道上运动时的速度变化。得出的结果是，在相同的时间里，太阳到地球的连线扫过的面积相等。推广后成为行星运动第二定律，即面积定律。

在推算火星轨道时，开普勒发现理论与观测之间有8'误差，改用椭圆轨道后则二者符合，于是得出结论：火星在椭圆轨道上绕太阳运动，太阳位于其中一个焦点上。推广后成为行星运动第一定律，即椭圆定律。

1609年，开普勒出版《新天文学》，公布了第一、二定律。

1609年至1619年期间，他继续研究，又发现行星到太阳的平均距离的立方与其公转周期的平方成正比，这是行星运动第三定律，又称调和定律。用公式表示为： $R_1^3/T_1^2=R_2^3/T_2^2$ 。

1619年，他又出版《宇宙的和谐》，公布了第三定律。这个行星运动的开普勒三定律为牛顿万有引力理论的发现起到了先导作用。

1617年至1621年间，他又分三部分出版了《哥白尼天文学概要》。

3. 其他成就

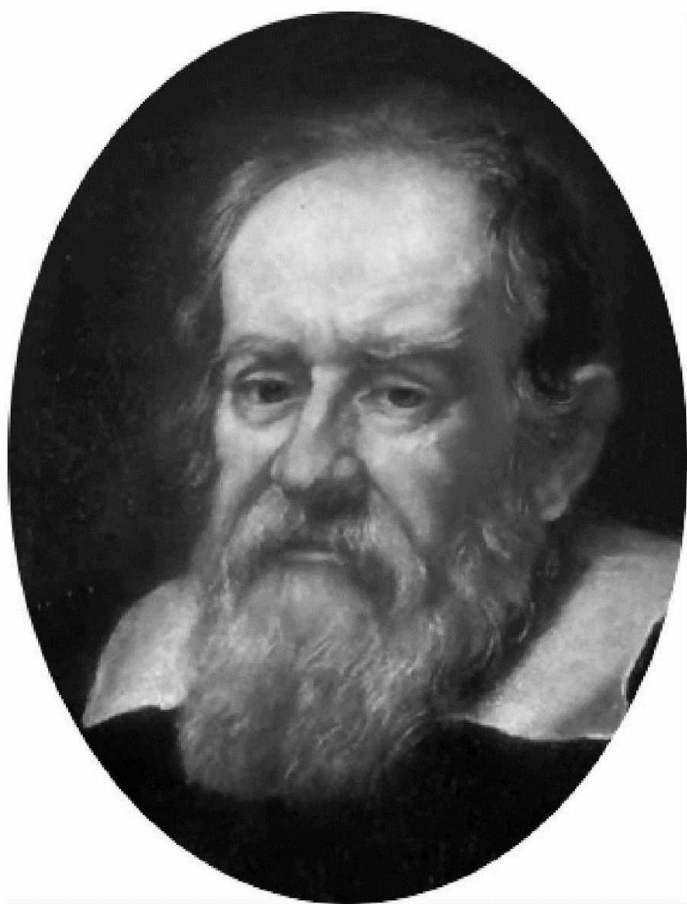
1604年发现了“开普勒新星”。

1611年出版了《折光学》。

1627年出版了《鲁道夫星表》。

他还探讨了行星绕太阳运动的物理原因，同时研究了大气折射、彗星运动和水星凌日。

伽利略·伽利雷（1564—1642）也是他们的同时代人，在帕多瓦大学执教时，他就读过哥白尼的著作《天体运行论》。但是伽利略是个科学态度十分严肃的学者。他想，过去都说是太阳围着地球运转，哥白尼却提出相反的看法，到底哪一个正确呢？伽利略没有轻率地下结论，他决定用自己的望远镜来证实谁是谁非。



伽利略



开普勒

当伽利略的著作《星际使者》出版时，他已是一个哥白尼学说的坚定支持者了。伽利略通过自己的观测和研究，逐渐认识到哥白尼的学说是正确的，而托勒密的地球中心说是错误的，亚里士多德的许多观点也是站不住脚的。伽利略不仅发表了批驳亚里士多德的论文，还通过书信毫不掩饰地支持哥白尼的学说，甚至把信件的副本直接寄给罗马教会。在伽利略看来，科学家的良心就是追随真理。

但是，罗马教廷是不会放过伽利略的。他们先是对伽利略发出措辞严厉的警告，后来又把他召到罗马进行审讯。1616年2月，宗教裁判所宣布，不许伽利略再宣传哥白尼的学说，无论是讲课或写作，都不得再把哥白尼学说说成是真理。在教会的威胁下，伽利略被迫作了放弃哥白尼学说的声明。他怀着极其痛苦的心情回到佛罗伦萨，在沉默中度过了

好些年。

但是伽利略的内心深处并没有放弃哥白尼学说，相反，继续不断地观测和深入研究，使他更加坚信哥白尼学说是正确的科学理论。在佛罗伦萨郊外的别墅里，伽利略过着与世隔绝的生活。他的身体大不如前，但却依然念念不忘宣传哥白尼的学说。经过长久的酝酿构思，用了差不多5年时间，一部伟大的著作《关于两种世界体系的对话》终于诞生了。

《关于两种世界体系的对话》表面上是以三个人对话的形式，客观地讨论托勒密的地心说与哥白尼的日心说，对谁是谁非进行没有偏见的探讨。但是当这本书在1632年2月出版时，细心的读者不难看出，这本书以充分的论据和大量无可争辩的事实，有力地批判了亚里士多德和托勒密的错误理论，科学地论证哥白尼的日心说，宣告了宗教神学的彻底破产。

很快，教会嗅出了这本书包含的可怕思想，从字里行间流露出来的大胆结论使神学家们感到极大恐慌。那些对伽利略心怀不满的学术骗子和教会勾结、罗织罪名、策划阴谋，为迫害伽利略大造舆论。科学和神学不可调和的斗争爆发了。1632年8月，罗马宗教裁判所下令禁止这本书出售，并且由罗马教皇指名组织一个专门委员会对这本书进行审查。伽利略也接到了宗教裁判所要他去罗马接受审讯的一纸公文。

但这时候的伽利略已是69岁的老人，且病魔缠身、行动不便，许多关心他的人为他说情，但是罗马教皇恼怒地说：“除非证明他不能行动，否则在必要时就给他戴上手铐押来罗马！”就这样，1633年初，伽利略抱病来到罗马。他一到罗马便失去自由，被关进了宗教裁判所的牢狱，并且不准任何人和他接触。

人类历史上一次骇人听闻的迫害就这样开始了。在罗马宗教裁判所充满血腥和恐怖的法庭上，真理遭到谬误的否决，科学受到神权的审判。那些教会法官们，用火刑威胁伽利略放弃自己的信仰。年迈多病的伽利略坚信：真理是不可能用暴力扑灭的。尽管他可以声明放弃哥白尼学说，但是宇宙天体之间的秩序是谁也无法更改的。在审讯和刑法的折磨下，他被迫在法庭上当众表示忏悔，并在判决书上签了字。被判终身监禁。

伽利略的晚年是非常悲惨的。1637年，他由于白内障的恶化双目完全失明，而他唯一的亲人——小女儿先他离开人间，又给了他很大的打击。但是，即使这样，伽利略仍旧没有失去探索真理的勇气。1638年，他的一部《关于两门新科学的讨论》在朋友帮助下在荷兰出版，这本书是伽利略长期对物理学研究的系统总结，也是现代物理的第一部伟大著作。

伽利略在科学上为人类做出了巨大贡献，是经典物理学的奠基人，被誉为“近代力学之父”、“现代科学之父”和“现代科学家的第一人”。他在力学领域进行过著名的比萨斜塔重物自由下落实验，推翻了亚里士多德关于“物体落下的速度与重量成正比例”的学说，建立了自由落体定律；发现物体的惯性定律、摆振动的等时性和抛体运动规律，并确定了伽利略相对性原理；发明了摆针和温度计；他还是利用望远镜观察天体取得大量成果的第一人，开创了望远镜天文学时代。他发现了月球表面凹凸不平、木星的四个卫星、太阳黑子，银河由无数恒星组成，以及金星、水星的盈亏等现象。

1642年1月8日，78岁的伽利略停止了呼吸，但是他毕生捍卫的科学真理却与世长存。具有讽刺意味的是，300多年后，1979年11月，在世界主教会议上，罗马教皇提出重新审理“伽利略案件”。为此，世界著名

科学家组成了一个审查委员会，负责重新审理这一冤案。其实，哪里还用得着审理什么呢？宇宙飞船在太空飞行、人类的足印深深地留在月球表面、人造卫星的上天，宇宙探测器飞出太阳系并发回电波……所有这些现代科学技术，早已把中世纪教廷的黑暗钉在了耻辱柱上。人们将永远记住伽利略的名字，为了纪念他的功绩，人们把木卫一、木卫二、木卫三和木卫四命名为“伽利略卫星”。

3.2.6 牛顿与经典宇宙学的创立

哥白尼科学革命的完成、经典宇宙学理论的确立，最终落在科学巨匠艾萨克·牛顿（1643—1727）肩上。1687年，牛顿发表了他的不朽巨著《自然哲学的数学原理》，提出了万有引力和三大运动定律，奠定了天体力学的基础，也创立了以万有引力为核心的经典宇宙学理论。这些定律成为此后几个世纪里物理世界的基本科学观点，并构成了现代工程学的基础。

牛顿在前人研究成果的基础上，依靠自己锲而不舍的努力、超凡的智慧所发现的万有引力定律是17世纪自然科学最伟大的成就之一。他通过论证开普勒行星运动定律与他的引力理论间的一致性，展示了地面物体与天体的运动都遵循着相同的自然定律；把地面上物体运动的规律和天体运动的规律统一了起来，对以后物理学和宇宙学的发展具有深远影响。它第一次解释了一种基本相互作用的规律，在人类认识自然的历史上树立了一



牛顿

座里程碑，为太阳中心说提供了强有力的理论支持，并推动了科学革命。

万有引力定律是描述物体之间都具有的相互作用引力的定律。定律内容如下：任意两个质点通过连心线方向上的力相互吸引，该引力的大小与它们的质量乘积成正比，与它们距离的平方成反比，与两物体的化学本质或物理状态以及中介物质无关，公式表示：
$$F = G \frac{M_1 M_2}{R^2}$$
； F ：两个物体之间的引力； G ：万有引力常数； M_1 ：物体1的质量； M_2 ：物体2的质量； R ：两个物体之间的距离。

实际上，伽利略在1632年已经提出了离心力和向心力的初步想法，而布里阿德在1645年就提出了引力平方比关系的思想。牛顿在1665年至1666年的手稿中，用自己的方式证明了离心力定律，1673年，惠更斯也独立地提出了圆周运动的离心力公式。

万有引力与相互作用物体的质量乘积成正比，是从发现引力平方反比定律过渡到发现万有引力定律的必要阶段。牛顿从1665年至1685年，花了整整20年的时间，才沿着离心力、向心力、重力、万有引力概念的演化顺序，提出“万有引力”这个概念。牛顿在《原理》第三卷中写道：“最后，如果由实验和天文学观测，普遍显示出地球周围的一切天体被地球重力所吸引，并且其重力与它们各自含有的物质质量成比例，则月球同样按照物质质量被地球重力所吸引。另一方面，它显示出，我们的海洋被月球重力所吸引；并且一切行星相互被重力所吸引，彗星同样被太阳的重力所吸引。由于这个规则，我们必须普遍承认，一切物体，不论是什么，都被赋予了相互的引力的原理。因为根据这个表象所得出的一切物体的万有引力的论证……”

牛顿在1665年至1666年间只使用了离心力定律和开普勒第三定律，

因而只能证明圆轨道上的引力平方反比关系。在1679年，他知道了运用开普勒第二定律，但在证明方法上没有突破，仍停留在1665年至1666年的水平。到1684年1月，哈雷、雷恩、胡克和牛顿不仅都能够证明圆轨道上的引力平方反比关系，而且都已经知道椭圆轨道上也遵守引力平方反比关系。但是，最后只有牛顿根据开普勒第三定律，从离心力定律演化出的向心力定律和数学上的极限概念（或微积分概念），用几何法证明了这个难题。

万有引力定律揭示了天体运动的规律，在天文学上和宇宙航行计算方面有着广泛的应用。它为实际的天文观测提供了一套计算方法，可以只凭少数观测资料，就能算出长周期运行的天体运动轨道。科学史上哈雷彗星、海王星、冥王星的发现，都是应用万有引力定律取得的重大成就。利用万有引力公式、开普勒第三定律等还可以计算太阳、地球等无法直接测量的天体的质量。牛顿还解释了月亮和太阳的万有引力引起的潮汐等现象。他指出潮汐的大小不但同月球的位相有关，而且同太阳的方位有关；他还用质点间的万有引力证明，密度呈球对称的球体对外的引力都可以用同质量的质点放在中心的位置来代替；牛顿预言地球不是正球体，岁差就是由于太阳对赤道突出部分的摄动造成的；他依据万有引力定律和其他力学定律，对地球两极呈扁平形状的原因和地轴复杂的运动，也成功地做了说明。他推翻了古代人类认为的神之引力。

在力学上，除了万有引力定律，牛顿还阐明了动量和角动量守恒的原理。在现代物理学中，更为普遍适用的是动量守恒定律、能量守恒定律与角动量守恒定律，它们既适用于微观和宏观，也应用于宇观；既应用于经典物理学，也应用于非经典物理学。它们的陈述都非常简明：“动量、能量、角动量既不可能凭空创造，也不可能凭空消失。”

动量守恒定律、能量守恒定律以及角动量守恒定律一起成为现代物

理学中的三大基本守恒定律。最初它们是牛顿定律的推论，但后来发现它们的适用范围远远广于牛顿定律，是比牛顿定律更基础的物理规律，是时空性质的反映，表述了物质运动的本质。牛顿在其中的贡献是不言而喻的。

在光学上，牛顿发明了反射望远镜。他曾致力于颜色现象和光本性的研究，并基于对三棱镜将白光发散成可见光谱的观察，发展出了颜色理论。1666年，他用三棱镜研究日光，得出的结论是，白光是由不同颜色（即不同波长）的光混合而成的，不同波长的光有不同的折射率。在可见光中，红光波长最长，折射率最小；紫光波长最短，折射率最大。牛顿的这一重要发现成为光谱分析的基础，揭示了光色的秘密。

牛顿还曾把一个磨得很精、曲率半径较大的凸透镜的凸面，压在一个十分光洁的平面玻璃上。在白光照射下可看到，中心的接触点是一个暗点，周围则是明暗相间的同心圆圈。后人把这一现象称为“牛顿环”。他还创立了光的“微粒说”，从一个侧面反映了光的运动性质，但他对光的“波动说”并不持反对态度。

1704年，牛顿著成《光学》，系统阐述他在光学方面的研究成果，其中他详述了光的粒子理论。他认为光是由非常微小的微粒组成的，而普通物质是由较粗的微粒组成的。并推测如果通过某种炼金术的转化，“难道物质和光不能互相转变吗？物质不可能从进入其结构中的光粒子得到主要的动力吗？”牛顿还使用玻璃球制造了原始形式的摩擦静电发电机。

他还系统地表述了冷却定律，并研究了音速。

在数学上，牛顿与德国学者戈特弗里德·威廉·莱布尼茨分享了创立

微积分学的荣誉。他也证明了广义二项式定理，提出了用以趋近函数零点的“牛顿法”，并为幂级数的研究做出了贡献。

此外，牛顿的哲学思想基本属于自发的唯物主义，他承认时间、空间的客观存在，但他也不能不受时代的局限。例如，他把时间、空间看作是同运动着的物质相脱离的东西，提出了所谓绝对时间和绝对空间的概念；他对那些暂时无法解释的自然现象归结为上帝的安排，提出一切行星都是在某种外来的“第一推动力”的作用下才开始运动的说法。

牛顿的巨作《自然哲学的数学原理》，开辟了大科学时代。他发现的运动三定律和万有引力定律，为近代物理学和力学奠定了基础，他的万有引力定律和哥白尼的日心说奠定了现代宇宙学的理论基础。直到今天，人造地球卫星、火箭、宇宙飞船的发射升空和运行轨道的计算，都仍以这作为理论根据。在2005年，英国皇家学会进行了一场“谁是科学史上最有影响力的人”的民意调查，牛顿被认为比阿尔伯特·爱因斯坦更具影响力。

3.2.7 万有引力定律的证实

1. 哈雷彗星回归的预言

1705年，哈雷发表《彗星天文学概论》，以万有引力定律计算了历史上24颗彗星的轨道。他发现1531、1607、1682年的彗星轨道几乎相同，于是成功预言1758年会有彗星回归。

2. 孤立大山的引力使铅垂线偏转

1774年，格林尼治天文台台长马斯基林（1732—1811）测量了苏格兰希哈里恩山的南北纬度差。天文观测值为 $54.6''$ ，大地测量为 $42.94''$ 。

这一结果证实了大山的引力使铅垂线发生偏转。

3. 万有引力常数的测定

1798年，英国物理学家卡文迪许（1731—1810）通过扭秤实验求得 G 。

4. 海王星的发现

1781年，英国天文学家威廉·赫歇尔（1738—1822）发现了天王星。其位置的理论值与观测值存在偏差。

1845年9月，亚当斯（1819—1892）算出了一颗未知行星的轨道，并向时任格林尼治天文台台长艾里（1801—1892）报告。

1846年8月，法国人勒威耶（1811—1877）发表“论使天王星失常的行星，它的质量、轨道和现在位置的决定”。9月，他写信给柏林天文台的伽勒（1812—1910），请他寻找这颗行星。9月23日，一颗新行星被发现，被命名为“海王星”。

5. 对恒星暗伴星的预言和发现

1834年，贝塞尔（1784—1846）发现天狼星的自行不是直线，预言其有伴星存在。1840年发现的南河三亦然。

1862年，克拉克（1832—1897）发现了天狼星的暗伴星。

1892年，舍伯尔（1853—1924）发现了南河三的暗伴星。

这些发现证实了万有引力定律同样适用于恒星世界。

3.2.8 17世纪的天文望远镜及其观测成就

1. 天文望远镜问世

(1) 望远镜的发明

1608年，荷兰眼镜商利帕席（约1570—1619）发明了第一台望远镜。

(2) 伽利略式天文望远镜

1609年，伽利略制作了一架口径0.42米，长约1.2米的望远镜。他用平凸透镜作为物镜，凹透镜作为目镜，采用这种光学系统的望远镜称为“伽利略式望远镜”。伽利略用这架望远镜看向天空，得到了一系列重要发现。

2. 伽利略的发现

(1) 观测月亮

1609年12月，伽利略通过自制的望远镜发现了“海”、山脉、环形山等，并画出了首张月面图。

(2) 观测恒星

观测到亮度较暗，数量更多的恒星。包括昴星团、银河中的众多恒星。

(3) 观测行星

1610年1月7日，伽利略开始用望远镜观看木星，发现4颗木星的卫

星。

1610年出版《星际使者》。1610年8月观看金星。

（4）观测太阳

1610年末开始观测太阳，发现太阳黑子和太阳自转。

1613年出版《关于太阳黑子的书信》。

伽利略的工作开创了望远镜天文学的时代。

3. 开普勒式望远镜

1611年，德国天文学家开普勒用两片双凸透镜分别作为物镜和目镜，使放大倍数有了明显提高，以后人们将这种光学系统的望远镜称为“开普勒式望远镜”。

需要指出的是，由于当时的望远镜采用单个透镜作为物镜，存在严重的色差，为了获得好的观测效果，需要用曲率非常小的透镜，这势必会造成镜身的加长。所以在很长的一段时间内，天文学家一直在梦想制作更长的望远镜，但许多尝试均以失败告终。

1757年，杜隆通过研究玻璃和水的折射和色散，建立了消色差透镜的理论基础，并用冕牌玻璃和火石玻璃制造了消色差透镜。从此，消色差折射望远镜完全取代了长镜身望远镜。但是，由于技术方面的限制，当时很难铸造较大的火石玻璃，在消色差望远镜的初期，最多只能磨制出口径0.1米的透镜。

19世纪末，随着制造技术的提高，制造较大口径的折射望远镜成为

可能，随之就出现了一个制造大口径折射望远镜的高潮。世界上现有的8架0.7米以上口径的折射望远镜中，有7架是在1885年到1897年期间建成的，其中最有代表性的是1897年在美国叶凯士天文台建成的口径1.02米望远镜，和1886年在德国里克天文台建成的口径0.91米望远镜。

折射望远镜的优点是焦距长、底片比例尺大，对镜筒弯曲不敏感，最适合于做天体测量方面的工作。但是它总是有残余的色差，同时对紫外、红外波段的辐射吸收很厉害。而巨大的



天文望远镜

的光学玻璃浇制也十分困难。并且，由于重力使大尺寸透镜的变形会非常明显，因而丧失明锐的焦点。到1897年叶凯士望远镜建成，折射望远镜的发展达到了顶点，此后的100年中再也没有更大的折射望远镜出现。

历史上的长焦距望远镜：1673年波兰天文学家赫维留斯（1611—1687）制成了46米长的望远镜。法国天文学家G.D.西尼（1625—1712）的望远镜长41.5米。荷兰物理学家兼天文学家惠更斯（1629—1695）的望远镜焦距37米。

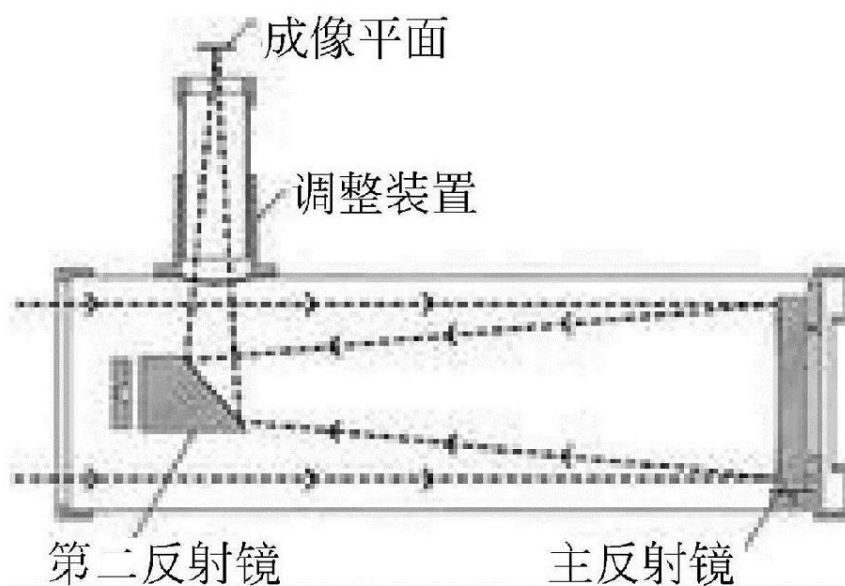
4. 牛顿式反射望远镜

1668年，牛顿制成第一架反射望远镜。他在经过多次研制非球面的透镜都不成功后，才决定用球面反射镜作为望远镜主镜。他把0.025米

直径的金属磨制成一个凹面反射镜，并在主镜的焦点前放了一个与主镜成 45° 角的反射镜，使经主镜反射后的会聚光经反射镜后以 90° 角反射出镜筒后，到达目镜。牛顿望远镜为反射望远镜的发展铺平了道路，现在的巨型望远镜中，大多属于反射望远镜。

牛顿反射望远镜采用抛物面镜作为主镜，光进入镜筒的底端，然后折回。

开口处的第二反射镜（平面的对角反射镜），使光线再次改变方向后进入目镜焦平面。目镜为便于观察，被安置靠近望远镜镜筒顶部的侧方。牛顿反射望远镜用平面镜替换了昂贵笨重的透镜收集和聚焦光线，从而增强了光线汇聚。



牛顿反射望远镜光学系统

牛顿反射望远镜系统的焦距长达1米，但仍然相对的紧凑和便携。不过因为主镜被暴露在空气和尘土中，这种望远镜要求更多地维护与保

养。

由于光学系统的原理，牛顿望远镜成倒像，但这并不影响天文观测，因此牛顿反射望远镜是天文学使用的最佳选择。通过正像镜等附加镜头，可以将图像校正过来，但会降低成像质量。

5. 天文望远镜的其他发现和成果

（1）赫维留斯和里乔利的月面图

1647年，赫维留斯出版专著《月面图》。意大利天文学家里乔利（1598—1671）描绘月面图。

（2）惠更斯的发现

1655年，惠更斯发现土卫六，1656年3月，发现土星光环。1659年出版《土星系》。

（3）卡西尼的发现

1664年，G.D.卡西尼发现木星带纹和大红斑。测定木星自转周期。观测木星凌日。1666年，测定火星的自转周期为24小时40分（今值为24小时37分）。1668年编制木卫星历表。1673年测太阳视差为9.5"，相应于日地距离1.4亿千米。1675年发现土星环缝，推测光环构成。

（4）罗默测定光速

丹麦天文学家罗默（1644—1710）于1672年起在巴黎天文台观测木星对木卫一的掩食。1676年，他宣布光穿过地球轨道直径需时22分钟（今测值16分38秒），由此算得光速约25万千米/秒，意义重大。

3.2.9 小结

一批又一批的科学勇士不畏艰险，胸怀追求真理的信念，冲破中世纪宗教神学的黑暗统治，迎来了科学的曙光。经典宇宙、天文学在以哥白尼和牛顿为代表的科技工作者卓有成效的努力下被确立起来。天文望远镜的发明，加强了天文观测的技术手段，极大地拓宽了人类的视野，获得了一系列科研成果，使宇宙学理论研究走上了必须用实验观测证实的科学轨道；也为宇宙、天文学的更大发展，奠定了坚实的基础。

参考文献

- [1] 陈久金. 中国古代天文学家 [M]. 北京: 中国科学技术出版社, 2008.
- [2] 沈括. 梦溪笔谈全译 [M]. 上海: 上海古籍出版社, 2013.
- [3] 王军云. 明朝二十四臣 [M]. 北京: 中国华侨出版社, 2007.
- [4] 李道英. 唐宋八大家的故事 [M]. 北京: 金盾出版社, 2013.
- [5] 黄坤. 朱熹诗文选译 [M]. 南京: 凤凰出版社, 2011.
- [6] 肖聿父. 许苏民. 王夫之评传 [M]. 南京: 南京大学出版社, 2007.
- [7] 曹国庆. 黄宗羲评传 [M]. 北京: 中国社会出版社, 2010.
- [8] 陈祖武. 顾炎武评传 [M]. 北京: 中国社会出版社, 2010.
- [9] 叶永烈. 钱学森 [M]. 上海: 上海交通大学出版社, 2010.
- [10] 百度网. 宗教神学 [OL]. 2015-06-29.
<http://baike.baidu.com/view/11760728.htm>
- [11] 中国科学技术大学天体物理组. 西方宇宙理论评述 [M]. 北京: 科学出版社, 1978.
- [12] [美] 阿若优. 生命四元素: 占星与心理学 [M]. 胡因梦, 译. 昆明: 云南出版集团公司, 2014.

- [13] 王元凯, 段金龙. 宇宙探索路 [M]. 北京: 科学技术文献出版社, 2008.
- [14] 《图说天下·世界历史系列》编委会. 图说天下·世界历史意大利 [M]. 长春: 吉林出版集团有限责任公司, 2009.
- [15] [法] 伏古勒尔. 天文学简史 [M]. 罗玉君, 译. 桂林: 广西师范大学出版社, 2003.
- [16] [波] 哥白尼. 天体运行论 [M]. 叶式辉, 译. 北京: 北京大学出版社, 2006.
- [17] [美] 杰安·若西. 逃亡与异端——布鲁诺传 [M]. 王伟, 译. 北京: 商务印书馆, 2014.
- [18] 钮卫星. 天文学史 [M]. 上海: 上海交通大学出版社, 2011.
- [19] [俄] 库兹涅佐夫. 伽利略传 [M]. 陈太先, 马世元, 译. 北京: 商务印书馆, 2001.
- [20] [美] 詹姆斯·格雷克. 牛顿传 [M]. 吴铮, 译. 北京: 高等教育出版社, 2004.
- [21] [英] 格里芬. 科学简史 [M]. 张帆, 译. 济南: 山东画报出版社, 2006.

4 天文观测新科技与宇宙、天文学新进展

4.1 天文观察和测试新科技

4.1.1 天文望远镜技术快速发展

17世纪早期的天文望远镜发挥了其巨大的作用。18、19世纪，尤其是20世纪航天科技的发展，实现了在太空中用天文望远镜进行观测。天文望远镜技术的惊人进步，对宇宙、天文学起到了极大的推动作用。

（1）施密特式折反射望远镜

1931年，德国光学家施密特用一块接近于平行板的非球面薄透镜作为改正镜，与球面反射镜配合，制成了可以消除球差和轴外像差的施密特式折反射望远镜。这种望远镜光力强、视场大、像差小，适合于拍摄大面积的天区照片，尤其是对暗弱星云的拍照效果非常突出。现在，施密特望远镜已经成了天文观测的重要工具。

（2）马克苏托夫式折反射望远镜

1940年，马克苏托夫用一个弯月形状的透镜作为改正透镜，制造出了另一种类型的折反射望远镜。它的两个表面是两个曲率不同的球面，它们相差不大，但曲率和厚度都很大。它的所有表面均为球面，比施密特式望远镜的改正板容易磨制，镜筒也比较短，但视场比施密特式望远镜小，对玻璃的要求也高一些。

由于折反射式望远镜能兼顾折射和反射两种望远镜的优点，非常适合业余的天文观测和天文摄影，所以得到了广大天文爱好者的喜爱。

望远镜的聚光能力随着口径的增大而增强。聚光能力越强，就能够看到更暗更远的天体，其实也就是看到了更早期的宇宙。所以天体物理的发展需要更大口径的望远镜。

但是，随着望远镜口径增大，一系列的技术问题也会接踵而来。例如，海尔望远镜的镜头自重达14 500千克，可动部分的重量为530 000千克，而5米镜更是重达800 000千克。一方面，望远镜的自重过大会使镜头变形相当明显。另一方面，镜体温度不均也会使镜面产生畸变，进而影响成像质量。而从制造方面看，传统方法制造望远镜的费用几乎与口径的平方或立方成正比，所以制造更大口径的望远镜必须另辟新径。

自20世纪70年代以来，望远镜的制造方面发展了许多新技术，涉及光学、力学、电子学、热学、信息、计算机、自动控制和精密机械等领域。这些技术使望远镜的制造突破了镜面口径的局限，并且降低造价、简化了望远镜结构；特别是主动光学和自适应光学技术的出现和应用，使望远镜的设计和制造有了一个飞跃。

从20世纪80年代开始，国际上掀起了制造新一代大型望远镜的热潮。优秀的传统望远镜中的卡塞格林焦点在最好的工作状态下，可以将80%的几何光能集中在0.6"范围内，而采用新技术制造的新一代大型望远镜可将80%的光能集中在0.2"~0.4"，甚至更好。

下面对几个有代表性的大型望远镜分别作一些介绍：

（1）凯克望远镜

凯克望远镜Keck I 和Keck II 分别在1991年和1996年建成，这是当时世界上已投入工作的最大口径的光学望远镜。这两台完全相同的望远镜都放置在夏威夷的莫纳克亚山上，将它们放在一起是为了做干涉观测。

它们的口径都是10米，由36块六角镜面拼接组成。每块镜面口径均为1.8米，而厚度仅为0.1米。通过主动光学支撑系统，可以使镜面保持极高的精度。在望远镜的焦面上有三个设备：近红外照相机、高分辨率CCD探测器和高色散光谱仪。

“像凯克这样的大望远镜，可以让我们沿着时间的长河，探寻宇宙的起源，还可以让我们看到宇宙最初诞生的时刻。”

（2）欧洲甚大光学望远镜

欧洲南方天文台自1986年开始，着手研制由4台8米口径望远镜组成的等效口径为16米的光学望远镜（Very Large Telescope, VLT）。这些望远镜排列在一条直线上。它们均为里奇·克莱琴（Richey-Chretien）光学系统，焦比是 $F/2$ ，采用地平装置。主镜采用主动光学系统支撑，指向精度为1"，跟踪精度为0.05"，镜筒重量为100 000千克，叉臂重量不到120 000千克。这4台望远镜可以组成一个干涉阵进行干涉观测，也可以单独使用。

（3）双子望远镜

双子望远镜（GEMINI）是以美国为主导的一个多国参与的望远镜，由美国大学天文联盟负责实施。它由两个8米口径的望远镜组成，一个放在北半球，一个放在南半球，以进行全天系统观测。其主镜采用主动光学控制，副镜作为倾斜镜以进行快速改正。它还将通过自适应光

学系统使红外区的观测波段接近衍射极限。

该工程于1993年9月开始启动，第一台于1998年7月在夏威夷开光，第二台于2000年9月于智利赛拉帕琼台址开光，整个系统在2001年验收后正式投入使用。

（4）昴星团光学 / 红外望远镜

这是一台8米口径的光学 / 红外望远镜（SUBARU）。它有三个特点：一是镜面薄，通过主动光学和自适应光学技术获得较高的成像质量；二是可实现0.1"的高精度跟踪；三是采用圆柱形观测室，自动控制通风和空气过滤器，使热湍流的排除达到最佳条件。此望远镜采用桁架式结构，可使主镜框与副镜框在移动中保持平行。由日本天文社团所属，位于美国夏威夷。

（5）大天区面积多目标光纤光谱天文望远镜（Large Sky Area Multi-Object Fibre Spectroscopic Telescope, LAMOST）。

这是中国已建成的有效通光口径为4米、焦距为20米、视场达20平方度的中星仪式反射施密特望远镜，又称为“郭守敬望远镜”。它的技术特色是：

①把主动光学技术应用在反射施密特系统，在跟踪天体运动中作实时球差改正，实现大口径和大视场兼备的功能。

②球面主镜和反射镜均采用拼接技术。

③多目标光纤（可达4000根，一般望远镜只有600根）的光谱技术是一个重要突破。

LAMOST把观测的极限星系星等降低到20.5等，比斯隆数字化巡天（Sloan Digital Sky Survey, SDSS）计划高2等左右，可以实现对107个星系的光谱普测，把观测目标的数量提高了1个量级。

（6）射电望远镜



射电望远镜

1932年，用无线电天线探测到来自银河系中心（人马座方向）的射电辐射，标志着人类打开了在传统光学波段之外进行观测的第一个窗口。

第二次世界大战结束后，射电望远镜的发明和使用带来了一系列重大的宇宙天文发现。

英国曼彻斯特大学于1946年建造了直径为66.5米的固定式抛物面射

电望远镜，1955年又建成了当时世界上最大的可转动式抛物面射电望远镜；到了20世纪60年代，美国在波多黎各的阿雷西博镇建造了直径达305米的抛物面射电望远镜。它是顺着山坡固定在地表面上的，不能转动。1974年改建，由38 774块铝板拼成球面。80年代又扩建为直径366米。

1962年，综合孔径射电望远镜的成功建造，实现了由多个较小天线组成后，获得相当于大口径单天线所能取得的效果。

1967年第一次记录到了甚长基线射电（very-long-baseline interferometry, VLBI）干涉条纹。

70年代，联邦德国在玻恩附近建造了100米口径的全向转动抛物面射电望远镜，这是世界上最大的可转动单天线射电望远镜。

80年代以来，欧洲的VLBI网（EVN）、美国的VLBA阵和日本的空间VLBI（VLBI Space Observatory Programme, VSOP）相继投入使用，这是新一代射电望远镜的代表，它们在灵敏度、分辨率和观测波段上都大大超过了以往的望远镜。

中国科学院上海天文台和乌鲁木齐天文站的两架25米口径射电望远镜作为正式成员，参加了美国的地球自转连续观测计划（continuous observation of the rotation of the earth, CORE）和欧洲的甚长基线干涉网（European VLBI network, EVN），这两个计划分别用于地球自转和高精度天体测量研究（CORE）和天体物理研究（EVN）。这种由各国射电望远镜联合进行甚长基线干涉观测的方式，起到了任何一个国家单独使用大望远镜都不能达到的效果。

另外，美国国家射电天文台（National Radio Astronomy

Observatory, NRAO) 研制的100米单天线望远镜 (Green Bank telescope, GBT), 采用了无遮挡 (偏馈)、主动光学等设计。该天线目前已经投入使用。

国际上将联合建造接收面积为1平方千米的低频射电望远镜阵 (square kilometre array, SKA), 该计划将使低频射电观测的灵敏度提高约两个量级。

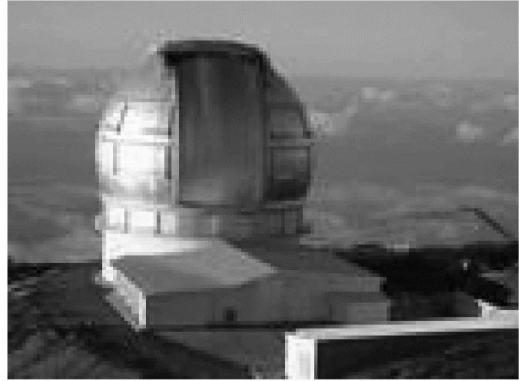
在增加射电观测波段覆盖方面, 美国史密松天体物理天文台和中国台湾天文与天体物理研究院正在夏威夷建造国际上第一个亚毫米波干涉阵 (Submillimeter array, SMA), 它由8个6米口径的天线组成, 工作频率从180吉赫到700吉赫, 部分设备已经安装。美国的毫米波阵 (millimeter array, MMA) 和欧洲的大南天阵 (large southern array, LSA) 将合并成为一个新的毫米波阵计划——阿塔卡马大型毫米 [1亚毫米] 波阵 (Atacama Large millimeter/Submillimeter array, ALMA)。这个计划将有64个12米天线组成, 最长基线达到10千米以上, 工作频率从70吉赫到950吉赫。

在提高射电观测的角分辨率方面, 新一代的大型设备大多数考虑采用干涉阵的方案。为了进一步提高空间VLBI观测的角分辨率和灵敏度, 第二代空间VLBI计划——高级空地射电干涉仪 (advanced radio interferometer between space and earth, ARISE) (25米口径) 已经提出。

相信这些设备的建成并投入使用将会使射电天文成为宇宙天文学的重要研究手段, 并会为宇宙天文学发展带来绝佳的发展机会。

2013年12月31日, 世界最大口径球面射电望远镜——500米口径球面射电望远镜 (five hundred meter aperture spherical radio telescope,

FAST) 在贵州省黔南布依族苗族自治州平塘县实现圈梁顺利合龙。该望远镜口径为500米, 占地约30个足球场大小。项目于2008年12月26日奠基, 预计2016年9月建成。

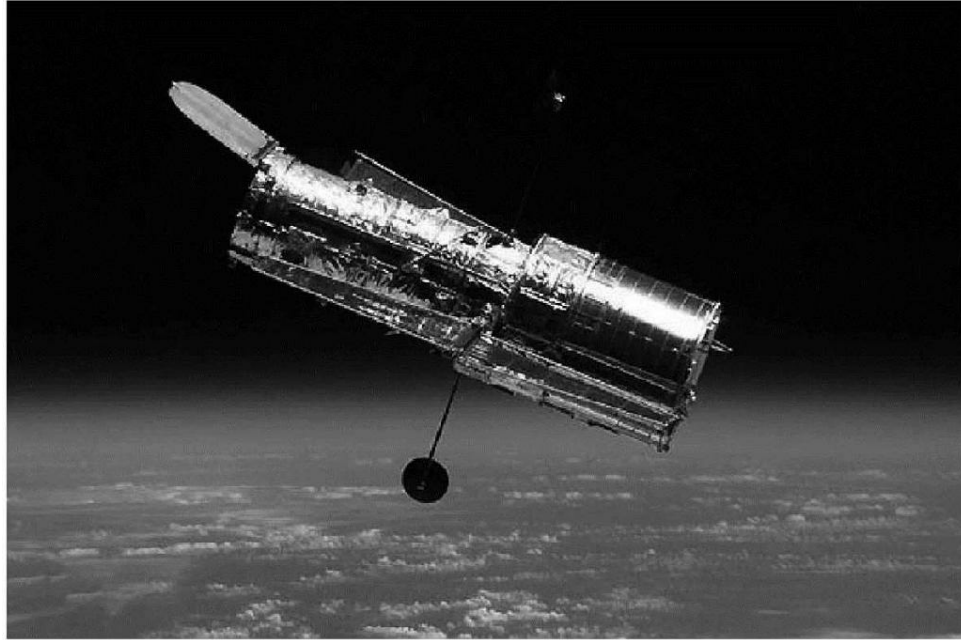


最大的红外天文望远镜

该望远镜建成后, 其口径将远超德国波恩的100米望远镜和美国波多黎各阿雷西博的300米望远镜, 将在未来20~30年保持世界一流设备的地位。

我们知道, 地球大气对电磁波有严重的吸收, 我们在地面上只能进行射电、可见光和部分红外波段的观测。随着空间技术的发展, 在大气外进行观测已成为可能, 所以就有了可以在大气层外观测的空间望远镜。空间观测设备与地面观测设备相比, 有极大的优势: 以光学望远镜为例, 望远镜可以接收到宽得多的波段, 短波甚至可以延伸到100纳米。没有大气抖动后, 望远镜的分辨本领得到很大提高。此外, 空间没有重力, 仪器也不会因自重而变形。后面介绍的紫外望远镜、X射线望远镜、Y射线望远镜以及部分红外望远镜的观测都是在地球大气层外进行的, 也属于空间望远镜。

(1) 哈勃空间望远镜



哈勃空间望远镜（Hubble space telescope, HST）是由美国宇航局主持建造的四座巨型空间天文台中的第一座，也是所有天文观测项目中规模最大、投资最多、最受到公众注目的一项。它筹建于1978年，设计历时7年，1989年完成，并于1990年4月25日由航天飞机运载升空，耗资30亿美元。但是由于人为原因造成了主镜光学系统的球差，所以人们不得不在1993年12月2日对其进行了规模浩大的修复工作。成功修复后的HST性能达到甚至超过了原先设计的目标。观测结果表明，它的分辨率比地面的大型望远镜高出几十倍。

在1997年进行的维修中，人们为HST安装了第二代仪器：有空间望远镜成像光谱仪、近红外照相机和多目标摄谱仪，把HST的观测范围扩展到了近红外，并提高了紫外光谱观测的效率。



绘画：张京

1999年12月的维修为HST更换了陀螺仪和新的计算机，并安装了第三代仪器——高级普查摄像仪，这将提高HST在紫外—光学—近红外的灵敏度和成像性能。

HST对宇宙天文学的发展起到了很重要的作用。

（2）其他空间天文望远镜

“新一代太空望远镜”（next generation space telescope, NGST）和“空间干涉测量任务”（space interferometry mission, SIM）是NASA“起源计划”的关键项目，用于探索在宇宙最早期形成的第一批星系和星团。其中，NGST是大孔径被动制冷望远镜，口径在4~8米之间，是HST和空间红外望远镜（space infrared telescope facility, SKTF）的后续项目。它强大的观测能力特别体现在光学、近红外和中红外的大视场、衍射成像方面。将运行于近地轨道的SIM采用迈克尔干涉方案，提供毫角秒级精度的恒星的精密绝对定位测量，同时由于其具有综合成像能力，能产生高分辨率的图像，所以可以用于实现搜索其他行星等科学目的。

“全天天体测量干涉仪”（global astrometric interferometer for astrophysics, GAIA）将会在对银河系的总体几何结构及其运动学做全面和彻底的普查，并在此基础上开辟广阔的天体物理研究领域。GAIA采用菲索干涉方案，视场为1°。GAIA和SIM的任务在很大程度上是互补的。

无人的空间天文观测只能依靠事先设计的观测模式自动进行，非常被动。如果能在月球表面建立月基天文台，那么就能化被动为主动，大

大提高观测精度。“阿波罗16号”登月时，宇航员在月面上拍摄的大麦哲伦星云照片表明，月面是理想的天文观测场所。建立月基天文台具有以下优点：

（1）月球上为高度真空状态，比空间天文观测设备所在位置还要低百万倍。

（2）月球为天文望远镜提供了一个稳定、坚固和巨大的观测平台，在月球上观测只需极简单的跟踪系统。

（3）月震活动只相当于地震活动的 10^{-8} ，这一点对于在月面上建立几十至数百公里的长基线射电、光学和红外干涉系统是很有利的。

（4）月球表面上的重力只有地球表面重力的1/6，这会给天文台的建造带来方便。另外，在地球上所有影响天文观测的因素，比如大气折射、散射和吸收，无线电干扰等，在月球上均不存在。

美国、欧洲和日本都计划在未来几年内再次登月，并在月球上建立永久居住区。可以预料，人类在月球上建立永久性基地后，建立月基天文台是可以期待的。

但是，对于宇宙、天文的科研领域来讲，空间观测项目是有局限性的。经费需求相当可观，如世界上最大的地面光学望远镜凯克的建设费用（7000万~9000万美元）要少于一颗普通的空间探测卫星的研制和发射费用。并且，空间天文观测的难度高、仪器的接收面积小、运行寿命短、难以维修，所以它并不能取代地面天文观测。在21世纪，空间观测与地面观测将是天文观测相辅相成的两翼。

我们知道，在地球表面有一层浓厚的大气，大气中各种粒子与天体

辐射的相互作用（主要是吸收和反射），使得大部分波段范围内的天体辐射无法到达地面。人们把能到达地面的波段形象地称为“大气窗口”，这种“窗口”有三个。

（1）光学窗口：这是最重要的一个窗口，波长在300~700纳米之间，包括了可见光波段（400~700纳米）。光学望远镜一直是地面天文观测的主要工具。

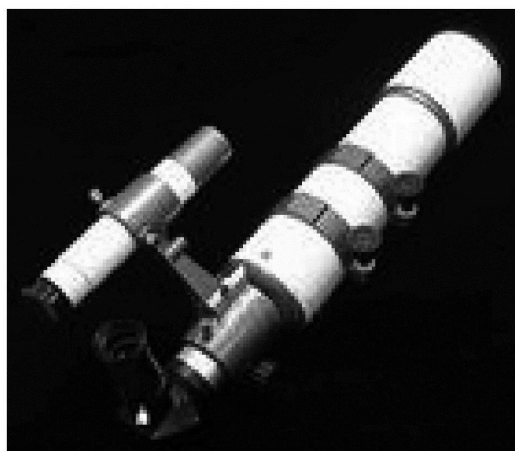
（2）红外窗口：红外波段的范围在0.7~1000微米之间，由于地球大气中不同分子吸收红外线波长不一致，造成红外波段的情况比较复杂。对于天文研究常用的有七个红外窗口。

（3）射电窗口：射电波段是指波长大于1毫米的电磁波。大气对射电波段也有少量的吸收，但在40毫米~30米的范围内，大气几乎是完全透明的。我们一般把1毫米~30米的范围称为射电窗口。

大气对于其他波段，比如紫外线、X射线、Y射线等均为不透明的，在人造卫星上天后才实现这些波段的天文观测。

（1）红外观测

最早的红外观测可以追溯到18世纪末。但是，由于地球大气的吸收和散射，在地面进行红外观测只能局限于几个近红外窗口。要获得更多红外波段的信息，就必须进行空间红外观测。现代的红外天文观测兴盛于19世纪60、70年代，当时是采用高空气球和飞机搭载红外望远镜或探测器进行



观测。

天文望远镜

1983年1月23日，美、英、荷联合发射了第一颗红外天文卫星（infrared astronomical satellite, IRAS）。其主体是一个口径为0.57米的望远镜，主要从事巡天工作。IRAS的成功极大地推动了红外天文在各个层次的发展。直到现在，IRAS的观测源仍然是天文学家研究的热点目标。

1995年11月17日，欧洲、美国和日本合作的红外空间天文台（infrared space observatory, ISO）发射升空并进入预定轨道。ISO的主体是一个口径为0.6米的RC式望远镜，它的功能和性能均比IRAS有许多提高。它携带了4台观测仪器，分别实现成像、偏振、分光、光栅分光、法布里—珀罗（F—P）干涉分光、测光等功能。与IRAS相比，ISO从近红外到远红外，有着更宽的波段范围；有更高的空间分辨率；更高的灵敏度（约为IRAS的100倍）；以及更多的功能。

ISO的实际工作寿命为30个月，用于对目标进行定点观测（IRAS的观测是巡天观测），这能有的放矢地解决天文学家提出的问题。预计在今后的几年中，以ISO数据为基础的研究将会成为天文学的热点之一。

红外望远镜与光学望远镜有许多相同或相似之处，因此可以对地面的光学望远镜进行一些改装，使它能同时也可从事红外观测。这样就可以用这些望远镜在月夜或白天进行红外观测，更大地发挥观测设备的效率。

（2）紫外波段观测

紫外波段是介于X射线和可见光之间的频率范围。在历史上紫外和可见光的划分界限在3900埃，当时的划分标准是肉眼能否看到。现代紫

外天文学的观测波段为3100~100埃，和X射线相接，臭氧层对电磁波的吸收界限在这里。紫外观测必须位于距离地表150千米的高度，以避开臭氧层和大气的吸收。第一次紫外观测是用气球将望远镜载上高空，后来用了火箭、航天飞机和卫星等航天技术，才使紫外观测有了真正的发展。

1968年美国发射了OAO轨道天文台—2（orbiting astronomical observatory-2, OAO-2），之后欧洲也发射了特德—1A（TD-1A），它们的任务是对天空的紫外辐射作一般性的普查观测。被命名为“哥白尼号”的OAO-3于1972年发射升空，它携带了一架0.8米口径的紫外望远镜，正常运行了9年，观测了波长在950~3500埃的天体紫外谱。

1978年发射了国际紫外探测者（international ultraviolet explorer, IUE），虽然其望远镜的口径比“哥白尼号”小，但检测灵敏度有了极大的提高。IUE的观测数据成为重要的天体物理研究资源。

1990年12月2日—11日，“哥伦比亚号”航天飞机搭载在“天星一号”紫外天文卫星（Astro-1）合作上，第一次实现了紫外光谱的天文观测；1995年3月2日开始，天星二号紫外天文卫星（Astro-2）完成了为期16天的紫外天文观测。

1992年美国宇航局发射了一颗观测卫星——极紫外探测器（extreme ultra-violet explore, EUVE），是在极远紫外波段作巡天观测。

1999年6月24日，远紫外探测器（far ultraviolet spectroscopic explorer, FUSE）卫星发射升空，这是NASA的“起源计划”项目之一，其任务是要回答天文学有关宇宙演化的基本问题。

紫外天文学是全波段天文学的重要组成部分，自“哥白尼号”升空至今的30年中，已经发展了紫外波段的极端紫外（EUV）、远紫外（FUV）、紫外（UV）等多种探测卫星，覆盖了全部紫外波段。

（3）X射线观测

X射线辐射的波段范围是0.01~10纳米，其中波长较短（能量较高）的称为“硬X射线”，波长较长的称为“软X射线”。天体的X射线是根本无法到达地面的，因此只有在60年代人造地球卫星上天后，天文学家才获得了重要的观测成果，X射线天文学才发展起来。早期主要是对太阳的X射线进行观测。

1962年6月，美国麻省理工学院的研究小组第一次发现来自天蝎座方向的强大X射线源，这使非太阳X射线天文学进入了较快的发展阶段。70年代，高能天文台1号（high energy astronomical observatory-1, HEAO-1）、2号（HEAO-2）两颗卫星发射成功，首次进行了X射线波段的巡天观测，使X射线的观测研究向前迈进了一大步。80年代以来，各国相继发射卫星，对X射线波段进行研究。

1987年4月，由苏联的火箭将德国、英国、苏联及荷兰等国家研制的X射线探测器送入太空。

1987年日本的“银河号”[X射线天文卫星]（Astro-C）发射升空。

1989年，苏联发射了一颗石榴号高能天文卫星（GRANAT），它载有苏联、法国、保加利亚和丹麦等国研制的7台探测仪器，主要工作为成像、光谱和对爆发现象的观测与监测。

1990年6月，伦琴X射线天文台（Röntgen satellite, ROSAT）进入地

球轨道，为研究工作取得大批重要的观测资料，目前它已基本完成预定的观测任务。

1990年12月，“哥伦比亚”号航天飞机将美国的“宽带X射线望远镜”（broad band X ray telescope, BBXRT）带入太空进行了为期9天的观测。

1993年2月，日本的“飞鸟”（Asuka）X射线探测卫星由火箭送入轨道。

1995年底美国发射了X射线时变探测器（X ray timing explorer, XTE）

1999年7月23日，美国成功发射了钱德拉X射线天文台（Chandra X ray observatory, CHANDRA）。

1999年12月13日，欧洲空间局发射了一颗名为X射线多镜望远镜（X ray multi-mirror mission, XMM）的卫星。

以上这些项目和计划表明，未来将会是一个X射线观测和研究的高潮。

（4） γ 射线观测

γ 射线比硬X射线的波长更短、能量更高。由于地球大气的吸收， γ 射线天文观测只能通过高空气球和人造卫星搭载的仪器进行。

1991年，美国的康普顿 γ 射线天文台（Compton γ ray observatory, CGRO）由航天飞机送入地球轨道。它的主要任务是进行 γ 波段的首次巡天观测，同时也对较强的宇宙 γ 射线源进行高灵敏度、高分辨率的成

像，进行能谱测量和光变测量，目前已取得了许多有重大科学价值的结果。

受到康普顿空间天文台成功的鼓舞，欧洲和美国的科研机构合作制订了一个新的 γ 射线望远镜计划——国际 γ 射线天体物理实验室（international γ ray astrophysics laboratory, INTEGRAL），在2002年送入太空，它的上天为康普顿空间天文台之后的 γ 射线天文学的进一步发展奠定了基础。

航天科技、空间天文台的使用，为宇宙天文学注入了新的活力，获得了许多新发现和新成果。



星云

这是位于美国亚利桑那州葛理翰山大学国际天文台天文望远镜（目前世界上最大的双目光学天文望远镜）拍到的第一张宇宙天体图片，这是一个距离地球1.02亿光年的螺旋形星系。

4.1.2 分光学、光度学、光谱学和照相术应用于宇宙、天文领域

光学望远镜精密度越来越高、口径越来越大，从而不断发现新天体、观测到新的天象。分光学、光度学、光谱学和照相术的出现，并应用于宇宙天文学领域，逐步奠定了太阳物理学、恒星物理学等天体物理学分支学科的基础。自从基尔霍夫说明了吸收线的产生原因以后，分光学在天体观测中就起了极重要的作用。通过观测和研究，人们不但能测定天体的温度、密度、压强等物理特性，而且还能得到天体化学成分的数据。太阳色球的单色光观测研究、太阳黑子磁场的发现、造父变星周光关系的发现、赫罗图的建立、星际消光的证明、星系是由恒星和星际物质组成的证明、星系的谱线红移以及银河系自转、恒星自转、星协、星链以至天王星光环的发现，都是光学天文学的重大成就。近几十年来射电天文学的兴起、红外天文学的复兴，以及紫外天文学、X射线天文学、 γ 射线天文学的诞生，使现代天体物理学进入自然科学的前沿阵地。光学天文学与上述各分支学科相互配合，仍然不断做出贡献，促进有关学科向前发展。

1. 太阳光的分解

1666年，牛顿用三棱镜分解白光，发现全色白光可分解为七色单色光。

1802年，英国物理学家沃拉斯顿（1766—1828）在棱镜前加狭缝观

测太阳，在彩带上发现暗线。

1814年，德国光学家夫琅禾费（1787—1826）制成第一台分光镜。他研究太阳光谱时，发现夫琅禾费线。

1853年，瑞典天文学家埃斯特罗姆（1814—1874）发现灼热气体会产生发射线，就是太阳光谱中的吸收线。

2. 光谱分析术的发明

1858年至1859年，德国化学家本生（1811—1899）与基尔霍夫（1824—1887）合作，发现了根据光谱判断元素的方法——光谱分析术。

3. 基尔霍夫定律的发现

基尔霍夫以实验探索太阳光谱中夫琅禾费线的本质，从而发现了基尔霍夫定律。这是认识太阳和恒星大气化学成分的准则。

4. 氦的发现

1868年8月18日，法国天文学家让桑（1824—1907）研究了日全食时日珥的光谱，发现一条橙黄色的明线。同年10月，英国天文学家洛基尔（1836—1920）也有同一发现。他们都报告给了法国科学院。

1869年，洛基尔认为此谱线来自一种新元素，命名为氦（helium，来源于希腊神话中太阳神Helios）。

1895年，英国化学家雷姆塞（1852—1916）在地球上发现这一元素。氦的发现证明了天体分光术的巨大成功。

5. 恒星光谱的分光观测

(1) 恒星光谱的观测和初步分类

1859年，英国的哈金斯（1824—1910）在0.2米口径望远镜上安装了高色散分光镜，用以观测亮星光谱。

1863年，意大利的塞奇（1818—1878）用低色散的分光镜观测了大量恒星。他于1868年提出一种光谱分类法，分类了4000颗恒星。

①白色星，光谱中只有几条氢的吸收线；

②黄色星，光谱同太阳光谱；

③橙色星和红色星，光谱里有明暗相间的暗带；

④暗红色星。

(2) 恒星光谱的谱线位移

1842年，奥地利物理学家多普勒（1803—1853）发现振动时振动波长会发生改变。

$$\Delta\lambda=\lambda\cdot V/V_s$$

式中， V_s 为声速或光速， V 为源运动速度， λ 为无运动振动波长。这被称为“多普勒效应”。

1868年，哈金斯根据恒星光谱的谱线位移测出了天狼星视向速度。视向速度的测量带来了许多重要的天文发现，如分光双星、新星和超新星的爆发、天体和天体系统的自转、可观测宇宙的膨胀等。

6. 天体测光术的发明和发展

(1) 恒星亮度的目视测量

公元前2世纪，古希腊天文学家依巴各首次估计出了恒星的视星等。

1782年，英国业余天文学家古德里克（1764—1786）发现变星大陵五（英仙座 β ）的光变周期。1785年和1786年，他又先后发现两颗变星造父一（仙王座 δ ）和渐台二（天琴座 β ）。由此开创了变星光度测量。

1852年至1859年，阿格兰德和匈费尔德（1828—1891）测量了恒星位置，并估计了星等，1863年发表了BD星表。1875年至1884年，匈费尔德测量了南天星空，并于1886年发表了SD星表。两星表共包含了457 874颗星。这是目视方法估计星等最宏伟的工作。

(2) 普森公式

19世纪上半叶，德国生理学家费希内尔（1807—1887）推出“感觉度随刺激度的对数变化”。

1856年，英国天文学家普森（1829—1891）建立了光度与星等间的基本关系式：

$$m_1 - m_2 = -2.5 \lg E_1/E_2。$$

普森公式为星等与光度之间的关系建立了定量关系，为科学的测光工作打下了基础。

(3) 目视光度计的发明

①偏振光度计。1859年，德国天文学家泽尔纳（1834—1882）发明了偏振光度计。

并于1861年发表了有266颗亮星的光度星表。

②光劈光度计。英国天文学家普里恰尔特（1808—1893）发明了光劈光度计。

并于1885年发表了有2784颗星的光度星表。

目视光度计的出现，从理论上和实测上开创了科学的天文光度学。

7. 天体照相术的应用

（1）照相术的发明和发展

1827年，法国艺术家尼普斯（1765—1833）拍出了人类历史上第一张照片。

19世纪30年代，法国艺术家达盖尔（1789—1851）发明用碘化银作底板的照相术。

后来，约翰·赫歇尔发明定影术。

1851年，英国摄影师斯科特-阿切尔（1813—1857）发明珂珞酊湿片法。

1871年，英国化学家马多克斯（1816—1902）发明明胶干板。

后来，有人提出底片敏化法。

德国化学家沃格尔（1834—1898）发明、扩展了光谱响应。

（2）照相术用于拍摄天体

1840年，美国化学家约翰·德雷珀（1811—1882）拍摄月亮，得到了世界上第一张天文照片，标志着天体摄影时代开始。

1896年，法国物理学家费佐（1819—1896）和傅科（1819—1868）在巴黎天文台首次拍摄太阳照片，照片上可见黑子。

1849年，美国天文学家威廉·邦德（1789—1859）拍出清晰的月亮照片（曝光20分钟）。

1850年，邦德与摄影家惠普尔合作，首次拍到了恒星——织女星。

1852年，天文学家德拉鲁（1815—1889）拍摄到了清晰的月亮（曝光30秒）。1860年拍到了日全食时的日珥照片。

1857年，惠普尔拍到了开阳（大熊座 δ ）和辅（开阳的伴星）的照片。

1882年，英国天文学家吉尔（1843—1914）拍到了大彗星照片，且恒星的像很清晰。

（3）照相术用于天体位置测量

1885年至1891年，吉尔拍摄了南天星空照片。1886年至1889年，荷兰天文学家卡普坦（1851—1922）测量了底片上恒星的位置。并于1896年至1900年发表了《好望角巡天星表》，刊载有454 875颗恒星（ $m > 10^m$ ， δ ： $-10^\circ \sim -90^\circ$ ）。

(4) 照相术用于拍摄天体光谱

1863年，哈金斯首次拍摄了恒星的连续光谱。

1872年，美国天文学家亨利·德雷珀拍摄织女星光谱，得到4条氫线。

19世纪末，哈佛大学天文台拍摄了大量恒星光谱，从而创立了哈佛光谱分类法。

照相法具有客观性、文献性、累积性的特点，因而得到了快速发展。

4.1.3 天体距离的测量

1. 恒星距离的测定

(1) 斯特鲁维的工作

俄国天文学家B.Я.斯特鲁维（1793—1864）于1836年用0.24米口径消色差望远镜对织女星（天琴座 α ）和近旁的一颗10.5等星作相对观测，测得视差为0.125″（今值0.121″）。

(2) 贝塞尔的工作

德国天文学家贝塞尔（1784—1846）选天鹅座61为观测目标，以两颗距离8'和12'的暗星为参照星进行测量。1838年发表视差值为0.314″（今值0.294″）。

(3) 亨德森的工作

英国天文学家亨德森（1798—1844），在1831年至1833年期间测量了半人马座 α 。1839年归算处理后，宣布视差为1.16"（今值0.76"）。

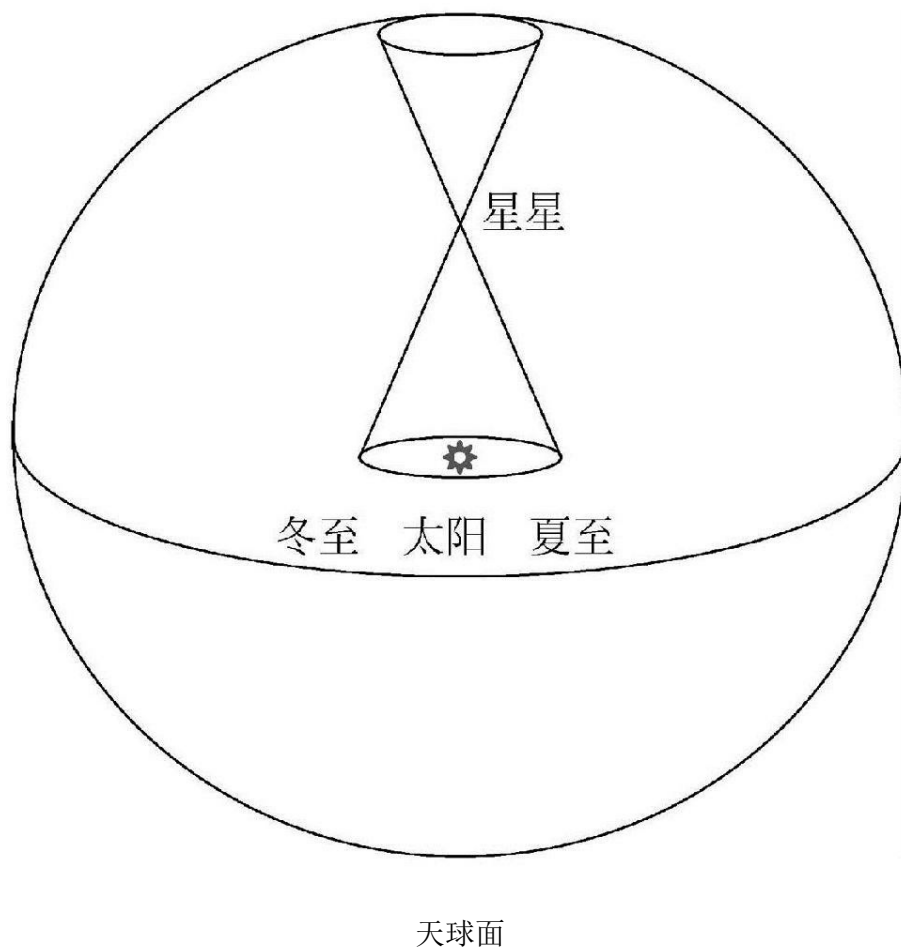
2. 天体距离的测量方法

（1）三角形法测量

在远离城市的旷野，我们在晴朗的夜空里可以看到满天的星星。可以发现，大多数星星的相互位置是不变的，天空就像是一张布满了亮点的大幕。这个大幕会慢慢地移动，它就叫做“天球面”，其示意图如下页图所示。

天球面并不是真实存在的，它只是看起来存在。这是因为，在地球绕着太阳公转的一年之中，距离大于300光年的星星在天球面上的位置看起来是不动的，于是这些星星看上去就形成了天球面。

实际上，也可以认为，天球面就是以太阳为球心、半径为300光年的一个球面。各种星座，如大熊星座、小熊星座.....中国的、外国的，都分布在天球面上。而距离小于300光年的星星看起来像是在天球面上画出一个椭圆轨迹。这个轨迹所张的最大的角叫做该星星的岁差。显然，有了岁差和日地距离，我们就可以计算得知这个星星的距离了。1弧秒的天空距离叫做“1秒差距”，1秒差距大约等于3光年。这种度量星星距离的方法叫做三角形法。



(2) 超远距离度量的阶梯

对于距离大于300光年的星星，岁差几乎为0。三角形法没有办法用了。人们只得找其他办法。对于遥远的星星，我们接触到的只有它的光。所以，只能从光着手。

晚上看书时，如果觉得暗了，可以移动得离灯近一些；或者，把灯移近些。这个常识告诉我们：灯的亮度与距离相关。星星也是这样，离我们近的，亮些；远的，暗些。问题是，必须要用同一种“灯”，同一种“灯泡”，测量恒星时类似，必须使用同一种、比较标准的星星。

星星有许多种，亮度差别很大。经过认真观测、选择，科学家发现

一些星星的亮度是变化的。例如，一种叫做“造父变星”的，其亮度变化是很有规律的。由造父变星的周光关系可以确定视差。

这样，根据这个规律，我们就可以找出许多造父变星。先用三角形法测量出比较近（小于300光年）的造父变星的亮度和距离。那么，测量其他造父变星的亮度，再与已经知道距离的造父变星的亮度比较，就可以得到其他造父变星的距离了。

这样，科学家一级一级地，就可以把测量的距离越扩越大。下表是通过这种方法得到的一些天体距离。

测量的部分天体距离

	天琴座 RR 型星	造父变星	新星	超巨星	超新星
最大距离/兆光年	0.6	6	13	50	320

还有测定恒星距离常使用一些间接的方法，如分光视差法、星团视差法、统计视差法等。这些间接的方法都是以三角视差法为基础的。自20世纪20年代以后，许多天文学家开展了这方面的工作，到20世纪90年代初，已有8000多颗恒星的距离被用照相方法测定。在20世纪90年代中期，依靠“依巴谷”卫星进行的空间天体测量获得成功，在大约3年的时间里，以非常高的准确度测定了10万颗恒星的距离。

4.2 经典宇宙理论的创立与发展

4.2.1 经典宇宙理论的力学和数学基础

1687年，牛顿出版巨著《自然哲学的数学原理》。

17世纪下半叶，牛顿和德国数学家莱布尼茨各自独立创立了微积分。

18世纪中叶，欧拉和拉格朗日等人创立了分析力学。

瑞士科学家欧拉（1707—1783）于1744年出版《行星和彗星的运动理论》，这是经典天体力学的第一部著作。



欧拉

1748年至1752年期间，欧拉在研究木星和土星的相互摄动时首创根数变易法，创立摄动理论。

行星运动=太阳引力作用（开普勒运动）+摄动→轨道根数变化

1753年，他提出第一个较完整的月球运动理论，后又提出改进的月球运动理论。

早在17世纪，牛顿就提出了力学的基本定律。欧拉特别擅长论证如何把这些定律运用到一些常见的物理现象中。例如，他把牛顿定律运用

到流体运动，建立了流体力学方程；同样，他通过认真分析刚体的可能运动，并应用牛顿定律建立了一个可以完全确定刚体运动的方程组。欧拉对弹性力学也做出了贡献。

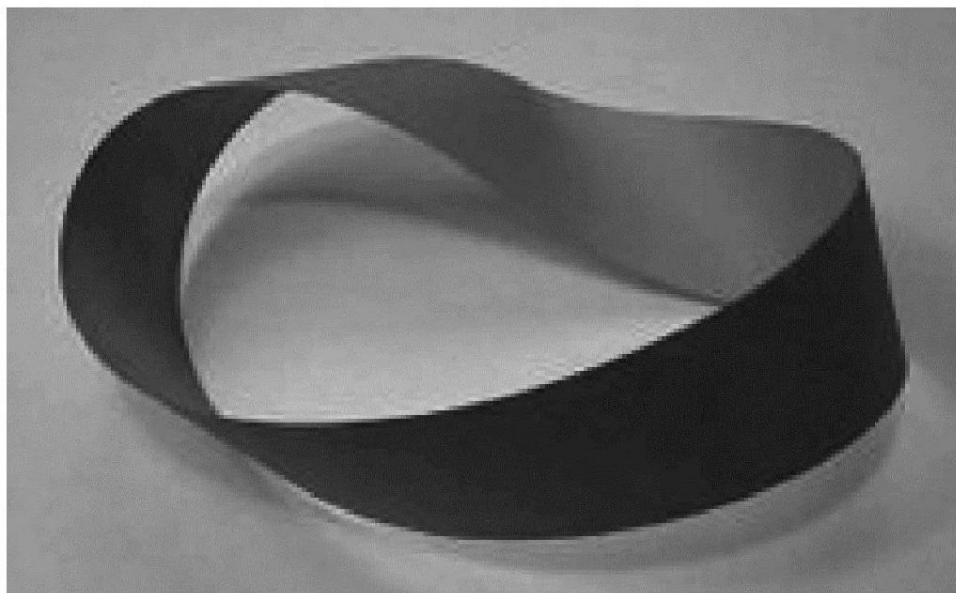
欧拉的天才还在于他用数学来分析天文学问题，特别是三体问题，即太阳、月亮和地球在相互引力作用下怎样运动的问题。这个问题——21世纪仍要面临的一个问题——尚未得到完全解决。顺便提一下，欧拉是18世纪独一无二的杰出科学家。他支持光波学说，结果证明他是正确的。

欧拉睿智的头脑常常为他人做出著名的发现开拓前进的道路。例如，法国数学家和物理学家约瑟夫·路易斯·拉格朗日创建了一个方程组，叫做“拉格朗日方程”。此方程在理论上非常重要，而且可以用来解决许多力学问题。但是由于基本方程是由欧拉首先提出的，因而通常称为“欧拉-拉格朗日方程”；人们一般认为另一名法国数学家让·巴普蒂斯·约瑟夫·傅里叶创造了一种重要的数学方法，叫做“傅里叶分析法”，但事实上其基本方程也是由欧拉最初创立的，因而叫做“欧拉—傅里叶方程”。这套方程在物理学的许多不同的领域都有着广泛的应用，其中包括声学 and 电磁学。

在数学方面，他对微积分的两个领域——微分方程和无穷级数——特别感兴趣。他在这两方面也做出了非常重要的贡献，他对变分学和复数学的贡献为后来所取得的一切成就奠定了基础。此外，欧拉还编写了一本解析几何的教科书，对微分几何和普通几何都做出了有意义的贡献。

欧拉不仅在做可应用于科学的数学发明上得心应手，而且在纯数学领域也具备几乎同样杰出的才能。他是数学的一个分支拓扑学领域的先

驱，拓扑学在20世纪就已经变得非常重要。



欧拉的贡献之一——拓扑学

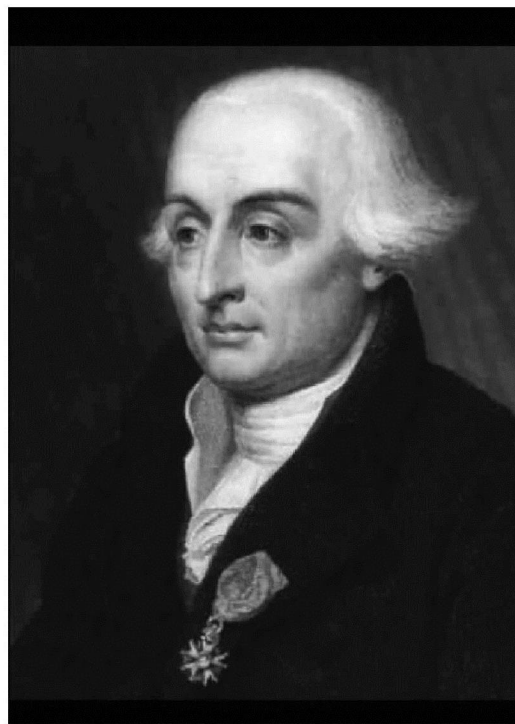
最后要提到的一点也很重要，欧拉对使用的数学符号制作出了重要的贡献。例如，常用的希腊字母 π 代表圆周率就是他提出来的。他还引出许多其他简便的符号，数学中经常使用到。

欧拉的著述浩瀚，不仅包含科学创见，而且富有科学思想，他给后人留下了极其丰富的科学遗产和为科学献身的精神。历史学家把欧拉同阿基米德、牛顿、高斯并列为数学史上的“四杰”。

拉格朗日（1736—1813）是法国科学家，1788年发表巨著《分析力学》。

他研究了太阳系的稳定性问题，1774年至1776年又讨论了行星轨道的半长轴 a 和偏心率 e 是否有长期摄动，证明 a 的变化是周期的。提出了计算长期摄动方法，并与拉普拉斯一起提出了在一阶摄动下的太阳系稳定性定理。此外，拉格朗日级数在摄动理论中也中有广泛应用。

天体力学是在牛顿发表万有引力定律（1687）时诞生的，很快就成为天文学的主流。它的学科内容和基本理论是在18世纪后期建立的。其主要奠基者为欧拉、A.C.克莱罗、达朗贝尔、拉格朗日和拉普拉斯。最后由拉普拉斯集大成而正式建立经典天体力学。拉格朗日一生的研究工作中，约有一半同天体力学有关。但他主要是数学家，他把力学作为数学分析的一个分支，而又把天体力学作为力学的一个分支对待。虽然如此，他在天体力学的奠基过程中，仍作出了重大历史性贡献。



拉格朗日

拉格朗日创立分析力学使力学发展到新的阶段。首先，在建立天体运动方程上，拉格朗日用他在分析力学中的原理和公式，建立起各类天体的运动方程。其中，特别是根据他在微分方程解法时所用的任意常数变异法，建立了以天体椭圆轨道根数为基本变量的运动方程，仍称作“拉格朗日行星运动方程”。它推广了牛顿第二运动定律，使得在任意坐标系下都有统一形式的运动方程，便于处理各种约束条件等优点，至今仍为动力学中的最重要的方程。此方程对摄动理论的建立和完善起了重大作用，得到达朗贝尔和拉普拉斯的高度评价。

在天体运动方程解法中，拉格朗日的重大历史性贡献是发现三体问题运动方程的五个特解，即拉格朗日平动解。其中两个解是三个天体围绕质量中心作椭圆运动过程中，永远保持等边三角形。他的这个理论结果在

100多年后得到证实。1907年2月22日，德国海德堡天文台发现了一颗小行星（后来命名为希腊神话中的大力士阿基里斯（Achilles），编号为588），它的位置正好与太阳、木星形成等边三角形。到1970年之前，人们已发现15颗这样的小行星，都以希腊神话中特洛伊（Troy）战争中将帅们的名字命名。有9颗位于木星轨道上前面 60° 处的拉格朗日特解附近，名为希腊人（Greek）群；有6颗位于木星轨道上后面 60° 处的解附近，名为特洛伊（Trojan）群。1970年以后，人们又陆续发现40多颗小行星位于此两群内，其中我国紫金山天文台发现4颗，但尚未命名。至于为什么在特解附近存在小行星，是因为这两个特解是稳定的。1961年，人们又在月球轨道前后发现与地月组成等边三角处聚集有流星物质，这是拉格朗日特解的又一证明。至今尚未找到肯定在三个拉格朗日共线群（三体共线情况）处附近的天体，因为这三个特解不稳定。

在具体的天体运动研究中，拉格朗日也有大量重要贡献，其中大部分是参加巴黎科学院征奖的课题。他的月球运动理论研究论文多次获奖。1763年完成的“月球天平动研究”获1764年度奖，此文较好地解释了月球自转和公转的角速度差异，但对月球赤道和轨道面的转动规律解释得不够好，后来在1780年完成的论文解决得更好；获1772年度奖的就是著名的三体问题论文，也是针对月球运动研究写出的；获1774年度奖的论文为“关于月球运动的长期差”，其中第一次讨论了地球形状和所有大行星对月球的摄动；关于行星和彗星运动的论文也有两次获奖；1776年度获奖的是他在1775年完成的三篇论文，其中讨论了行星轨道交点和倾角的长期变化对彗星运动的影响；1780年度的获奖论文就是提出著名的拉格朗日行星运动方程的那篇；获1766年度奖的论文是“木星的卫星运动的偏差研究……”，其中第一次讨论了太阳引力对木星的4个卫星运动的影响，结果比达朗贝尔的更好。

拉格朗日从事的天体力学课题还有很多，如在柏林时期的前半部分，还研究了用3个时刻的观测资料计算彗星轨道的方法，所得结果成为轨道计算的基础。另外他还得到了一种力学模型——两个不动中心问题的解，这是欧拉已讨论过的，又称为“欧拉问题”。拉格朗日将其推广到存在离心力的情况，故后来又称为“拉格朗日问题”。这些模型现今仍在应用，例如有人将其用作人造卫星运动的近似力学模型。此外，他在《分析力学》中给出的流体静力学的结果，后来又成为讨论天体形状理论的基础。

总的来看，拉格朗日在天体力学的五个奠基者中，所做的历史性贡献仅次于拉普拉斯。他创立的“分析力学”对以后宇宙学的发展有深远的影响。

拉格朗日是18世纪的伟大科学家，在数学、力学和天文学三个学科中都有历史性的重大贡献。他的努力促使宇宙学得到了更深入的发展。但由于历史的局限，严密性不够仍然妨碍着他取得更多的成果。

皮埃尔-西蒙·拉普拉斯（1749—1827），法国科学家。是天体力学的主要奠基人、天体演化学的创立者之一，他还是分析概率论的创始人。



拉普拉斯

1773年，他证明木星和土星的轨道具有周期性，并给出定理。

1799年至1825年出版《天体力学》，共5卷16册，是天体力学的奠

基工作。

拉普拉斯把注意力主要集中在天体力学的研究上面。他把牛顿的万有引力定律应用到整个太阳系。1773年，他解决了一个当时著名的难题：木星轨道为什么在不断地收缩，而同时土星的轨道又在不断地膨胀。拉普拉斯用数学方法证明行星平均运动的不变性，即行星的轨道大小只有周期性变化，并证明为偏心率和倾角的3次幂。这就是著名的拉普拉斯定理。此后他开始对太阳系稳定性问题的研究。1784年至1785年，他求得天体对其外任一质点的引力分量可以用一个势函数来表示，这个势函数满足一个偏微分方程，即著名的拉普拉斯方程。

1786年，他证明行星轨道的偏心率和倾角总保持很小和恒定，能自动调整，即摄动效应是守恒和周期性的，不会积累也不会消解。拉普拉斯注意到木星的三个主要卫星的平均运动 Z_1 、 Z_2 、 Z_3 服从下列关系式： $Z_1 - 3 \times Z_2 + 2 \times Z_3 = 0$ 。同样，土星的4个卫星的平均运动 Y_1 、 Y_2 、 Y_3 、 Y_4 也具有类似的关系： $5 \times Y_1 - 10 \times Y_2 + Y_3 + 4 \times Y_4 = 0$ 。后人称这些卫星之间存在可公度性，并由此演变出时间之窗的概念。

1787年，他发现月球的加速度同地球轨道的偏心率有关，从理论上解决了太阳系动态中观测到的最后一个反常问题。

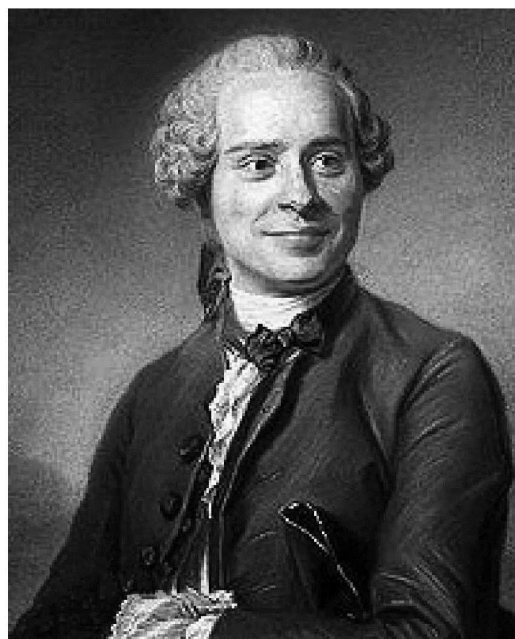
1796年，他的著作《宇宙体系论》问世，在这部书中，他独立于康德，提出了第一个科学的太阳系起源理论——星云说。对后世影响巨大。

他长期从事大行星运动理论和月球运动理论方面的研究，尤其是特别注意研究太阳系天体摄动、太阳系的普遍稳定性问题以及太阳系稳定性的动力学问题。他在总结前人研究的基础上取得大量重要成果，这些

成果集中在1799年至1825年出版的5卷16册巨著《天体力学》之内。这部著作第一次提出“天体力学”这一名词，是经典天体力学的代表作。他也因此被誉为法国的“牛顿”和“天体力学之父”。1814年，拉普拉斯还提出一个科学假设，假定如果有一个智能生物能确定，从最大天体到最轻原子的运动的现时状态，就能按照力学规律推算出整个宇宙过去状态和未来状态。后人把他所假定的智能生物称为拉普拉斯妖。

法国科学家达朗贝尔（1717—1783）于1749年用天体力学方法建立岁差、章动理论。

达朗贝尔在数学、力学和天文学等许多领域都做出了贡献。1746年，他与当时著名的哲学家狄德罗一起编纂了法国《百科全书》，并负责撰写数学与自然科学条目，是法国百科全书派的主要领导者。在百科全书的序言中，达朗贝尔表达了自己坚持唯物主义观点、正确分析科学问题的思想。在这一段时间之内，他还在心理学、哲学、音乐、法学和宗教文学等方面都发表了作品。



达朗贝尔

《动力学》是达朗贝尔最伟大的物理学著作。在这部书里，他提出了三大运动定律。第一运动定律给出惯性定律的几何证明；第二定律是用数学证明力的分析中的平行四边形法则的；第三定律是用动量守恒来表示的平衡定律。他还提出了达朗贝尔原理，它与牛顿第二定律相似，只是进行了移项。但这是概念上的变化，它的发展在于可以把动力学问题转化为静力学问题处理，这种动静法的观点对力学的发展产生了积极

的影响。还可以用平面静力的方法分析刚体的平面运动。这一原理使一些力学问题的分析简单化，而且为分析力学的创立打下了基础。达朗贝尔原理阐明，对于任意物理系统，所有惯性力或施加的外力，经过符合约束条件的虚位移，所做的虚功的总和等于零。作用于一个物体的外力与动力的反作用之和等于零。即

$$F + (-Ma) + N = 0 \quad (4-1)$$

其中 M ， a 为物体质量和加速度， F 为物体受到的直接外力， N 为物体受到的约束反作用力（也是外力）。

在没有约束时，相应的 $N=0$ ，式（4-1）成为

$$F - Ma = 0 \quad (4-2)$$

$$\sum (F_i - m_i a_i) \cdot \delta r_i = 0$$

达朗贝尔原理简化公式。

研究有约束的质点系动力学问题的一个原理，由达朗贝尔于1743年提出而得名。对于质点系内任一个质点，此原理的表达式为 $F+N-ma=0$ ，式中 F 为作用于质量为 m 的某一质点上的主动力， N 为质点系作用于质点的约束力， a 为该质点的加速度。从形式上看，上式与从牛顿运动方程 $F+N=ma$ 中把 ma 移项所得结果相同。于是，后人把 $-ma$ 看作惯性力，而把达朗贝尔原理表述为：在质点受力运动的任何时刻，作用于质点的主动力、约束力和惯性力互相平衡。

达朗贝尔对当时运动量度的争论提出了自己的看法，他认为两种量度是等价的，并模糊地提出了物体动量的变化与力的作用时间有关。1752年，达朗贝尔第一次用微分方程表示场，同时提出了关于流体力学的一个原理，虽然存在一些问题，但是达朗贝尔第一次提出了“流体速

度”和“加速度分量”的概念。达朗贝尔的力学知识为天文学领域做出了重要贡献。同时达朗贝尔发现了流体自转时平衡形式的一般结果，形成关于地球形状和自转的理论。他还发表了关于春分点的论文。

1760年以后，达朗贝尔继续进行他的科学研究。随着研究成果的不断涌现，达朗贝尔的声誉也不断提高，而且尤其以写论文快速而闻名。达朗贝尔是位多产科学家，他对力学、数学和天文学的大量课题都进行了研究。

法国科学家克雷洛（1713—1765）在月球运动理论的创立和天体形状和自转理论的创立上都有重要贡献。

德国科学家高斯（1777—1855）率先开展了小行星运动研究，他创立用3次观测决定天体运动轨道的计算方法。

1794年，他发明最小二乘法后便应用于轨道计算。1809年出版《天体按照圆锥曲线运动的理论》。

17世纪末到19世纪上半叶，天体力学诞生并取得重要成就。使天文学从单纯描述天体的视位置和几何关系进入到研究天体的相互作用，即从单纯的天体状态进入到研究天体运动原因。这是人类认识宇宙的一次重大飞跃。可以认为，经典宇宙学就此创立。

4.2.2 太阳系起源说及康德和拉普拉斯的星云说

1. 早期的太阳系起源说和形而上学的自然观

（1）笛卡儿的太阳系起源的涡动说

1644年法国科学家笛卡儿（1596—1650）在其《哲学原理》一书中提出太阳系起源的涡动学说。此学说在思想上具有进步性，但在学术上有局限性。

（2）牛顿关于太阳系起源的考虑

1692年12月和1693年1月，牛顿在致本特利主教的两封信中提出关于太阳、恒星和行星形成和开始运动的设想。他认为是上帝的有意设计和给予的“第一次推动”。

（3）布封的太阳系形成学说

1745年法国科学家布封（1707—1788）提出太阳系形成学说，认为曾经有一个彗星掠碰太阳，形成了太阳系。

（4）17至18世纪形而上学自然观的特点

自然界绝对不变。天文学上、生物学上都是永恒不变的。哲学上的目的论。宇宙不变论，最后必然导致神学。

2. 康德的太阳系星云假说

康德（1724—1804），德国哲学家。

1754年，康德发表了论文《论地球自转是否变化和地球是否要衰老》，对“宇宙不变论”大胆提出怀疑。他提出潮汐摩擦导致地球自转变慢。他的思想与形而上学思想对立，体现了天体的发展变化。

1755年，他发表《自然通史和天体论》（又译《宇宙发展史概论》）。首先提出太阳系起源的星云说。康德在书中指出：太阳系是由

一团星云演变来的。这团原始星云是由大小不等的固体微粒组成的，“天体在吸引最强的地方开始形成”，万有引力使得微粒相互接近，大微粒把小微粒吸引过去，凝成较大的团块。团块越来越大，引力最强的中心部分吸引的物质最多，先形成太阳。外面的微粒在太阳吸引下向中心体下落时，与其他微粒碰撞而改变方向，变成绕太阳的圆周运动。这些绕太阳运动的微粒又逐渐形成几个引力中心，这些引力中心最后凝聚成朝同一方向转动的行星。卫星形成的过程与行星类似。彗星则是在原始星云的外围形



康德

成，太阳对它们的引力较弱，所以彗星轨道的倾角多种多样。行星的自转是由于落在行星上的质点的撞击而产生的。康德还用行星区范围的大小来解释行星的质量分布（当时人们仅知水星、金星、地球、火星、木星、土星六颗大行星，十颗卫星和三十多颗彗星）。

康德星云说否定了牛顿的神秘的“第一推动力”和神创论，表达了朴素辩证法的观点：

- （1）物质的必然的、自己运动的观点；
- （2）引力相互作用的观点；
- （3）宇宙在空间和时间上无限性的观点；

(4) 事物的发生、发展和灭亡的普遍规律的观点；

(5) 人类是物质发展到一定阶段上的产物的观点。

这是自然观的一场革命。第一次提出了天体和自然界是不断发展的辩证观点，因而在形而上学的僵化的自然观上打开了第一个缺口，这是从哥白尼以来天文学取得的最大进步。

3. 拉普拉斯的太阳系起源的星云说

法国科学家拉普拉斯在1796年出版《宇宙体系论》提出太阳系起源的星云说。

拉普拉斯认为，形成太阳系的云是一团巨大、灼热、转动着的气体，大致呈球状。由于冷却，星云逐渐收缩。因为角动量守恒，收缩使转动速度加快，在中心引力和离心力的共同作用下，星云逐渐变为扁平的盘状。在星云收缩过程中，每当离心力与引力相等时，就有部分物质留下来，演化为一个绕中心转动的环，以后又陆续形成好几个环。这样，星云的中心部分凝聚成太阳，各个环则凝聚成各个行星。较大的行星在凝聚过程中同样能分出一些气体物质环来形成卫星系统。

比康德的假说更进步之处在于：

(1) 运用角动量守恒原理，避免斥力概念；

(2) 提出温度变化，为天体演化开创新起点；

(3) 排斥上帝的作用。

康德的学说侧重于哲理，而拉普拉斯则从数学和力学上进行论述。

因此，人们常将他们的学说称为“康德—拉普拉斯星云说。”拉普拉斯的科学论述加上他在学术界的威望，使星云说在19世纪被人们普遍接受。但由于科学发展水平的限制，这两种星云学说也有不少缺点和错误，曾一度被人们摒弃。不过目前不少天文学家认为，星云说的基本思想还是正确的。

4. 星云说的历史意义

突破了形而上学的宇宙观。

在天文学上开创了一个新领域——天体演化学，这是经典宇宙学的重要组成部分。

4.2.3 银河系概念的初步确立

关于恒星系统和银河系认识的演进过程：

古希腊学者早在公元前就提出过关于“恒星天”的观念。

15世纪中叶的古萨的尼古拉大主教、16世纪下叶的英国学者迪格斯、意大利学者布鲁诺等人都认为天上的恒星多得数不清。

17世纪初，意大利科学家伽利略首次用望远镜观察恒星。

1717年，英国天文学家哈雷把天狼（大犬座 α ）、大角（牧夫座 α ）、南河三（小犬座 α ）等恒星的位置与古希腊测定的位置比较，发现了恒星自行。

1727年，英国天文学家布拉德雷（1693—1762）发现光行差，在之后的测量中发现恒星与地球的距离应大于6至8光年。

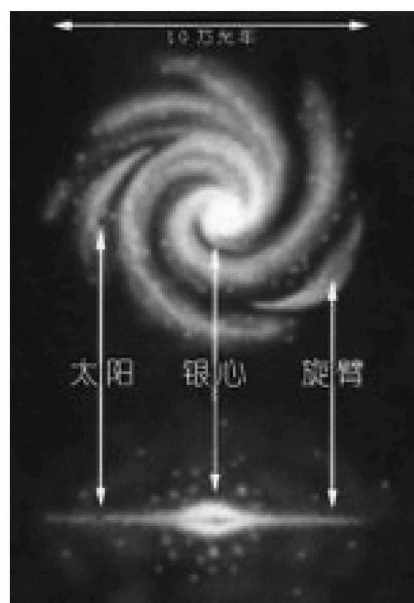
18世纪20至30年代，瑞典哲学家斯维登堡（1688—1772）最早推测银河系是宇宙中完整的力学体系。

1750年，英国天文学家赖特认为银河系是扁平的。他在《新颖的宇宙理论或新宇宙假设》一书中提出银河系概念及其形状。

1755年，德国哲学家康德提出了恒星和银河之间可能会组成一个巨大的天体系统；随后的德国数学家朗伯特也提出了类似的假设。这时，人们已意识到，除行星、月球等太阳系天体外，满天星斗都是远方的“太阳”。赖特、康德和朗伯特最先认为，很可能是全部恒星集合成了一个空间上有限的巨大系统。像太阳一样的恒星在银河系里是多之又多的！

1761年，德国学者朗伯（1728—1777）出版《宇宙论书简》，对于恒星世界结构，提出无限阶梯式的宇宙模型。

1785年，英国天文学家威廉·赫歇尔第一个通过观测研究恒星系统本原。他用自己的反射望远镜，共作了1083次观测，一共记录了683个取样天区中的117 600颗恒星。用“数星星”的方法绘制了一张银河图。在赫歇尔的银河图里，银河系是扁平的，群星环绕，其长度为7000光年，宽1400光年。我们的太阳处在银河系的中心，这是人类建立的第一个银河系结构图。它虽然很不完善，并且有错误，但使人类的视野从太阳系扩展到银河系广袤的恒星世界中。



威廉·赫歇尔持续用0.5米和1.2米口径望远镜观测，发现望远镜贯穿本领增加时，观察到的暗星也增多，但是仍然看不到银河系的边缘。他意识到，银河系远比他最初估计的大。赫歇尔死后，其子约翰·赫歇尔继承父业，将恒星的计数工作范围扩展到南半天。19世纪中叶，开始测定恒星的距离，并编制全天星图。

1783年威廉·赫歇尔考察了7颗亮星：天狼（大犬座 α ）、北河二（双子座 α ）、北河三（双子座 β ）、南河三（小犬座 α ）、轩辕十四（狮子座 γ ）、大角（牧夫座 α ）、河鼓二（天鹰座 α ）的自行。发现太阳向武仙座方向的本动。同年又用14颗恒星的自行，求出太阳本动方向在武仙座 λ 附近空间。

1837年，德国天文学家阿格兰德（1799—1875）用390颗恒星自行的观测获类似结果。

1834年至1838年，约翰·赫歇尔观测双星、星团和星云，同时统计了3000个选区的68 948颗恒星，证实了其父结论。多年后，到南非好望角建观测站，开拓了在南天的工作。

1849年，约翰·赫歇尔出版《天文学纲要》。1859年（咸丰九年），我国天文学家李善兰与传教士伟烈亚力合作翻译出版，译名《谈天》。近代天文学传入我国。

1845年，罗斯勋爵发现第一个旋涡星系M51。

1852年，美国天文学家史蒂芬·亚历山大声称银河系是一个旋涡星系，却拿不出证据加以证明。1869年，英国天文学作家理查·普洛托克提出相同的见解，但一样无法证实。

1906年，卡普坦为了重新研究恒星世界的结构，提出了“选择星区”计划，后人称为“卡普坦选区”。他于1922年得出与赫歇尔的类似的模型，也是一个扁平系统，太阳居中，中心的恒星密集，边缘稀疏。

1918年，美国天文学家沙普利在完全不同的基础上，探讨了银河系的大小和形状。他利用1908年至1912年勒维特发现的麦哲伦云中造父变星的周光关系，测定了当时已发现有造父变星的球状星团的距离。在假设没有明显星际消光的前提下，经过4年的观测，建立了银河系透镜模型，提出太阳系应该位于银河系的边缘，而不是在中心。到20世纪20年代，沙普利模型已得到天文界公认。由于未计入星际消光效应，沙普利把银河系估计得过大了。到1930年，特朗普勒证实星际物质存在后，这一偏差才得到纠正。

1926年，瑞典天文学家林得布拉德分析出银河系也在自转。

银河系概念的初步确立是人类对宇宙认识史上一个重大里程碑。从此，人们的眼界从狭小的太阳系扩展到浩瀚的恒星世界，视野大为开阔。是进一步认识整个宇宙的一个阶梯。

4.2.4 恒星天文学之父——英国天文学家威廉·赫歇尔

从前面所讲，可以看到英国天文学家威廉·赫歇尔（1738—1822）父子所作的杰出贡献，下面继续讲述威廉·赫歇尔这位天文学史上的传奇人物。

1781年，太阳系的第7颗大行星——天王星的发现，彻底改变了人类对太阳系的认识。发现者威廉·赫歇尔从此蜚声天下，从一个爱好天

文学的乐师变成了精通乐理的天文学家。

为了纪念200年前这一划时代的发现，英国格林尼茨海军大学于1981年4月25日举办了一场别开生面的“纪念赫歇尔音乐演奏会”。当时礼堂前车水马龙，会场上春意盎然。而那天演的所有节目，不论是交响乐还是奏鸣曲，也无论是协奏曲还是田园诗，全都是当年赫歇尔创作的作品。所有与会的音乐家和天文学家，众口称赞赫歇尔是世上少有的“音乐界和天文学界的双星”。的确，赫歇尔正是双



威廉·赫歇尔

星研究的奠基人，他证实了太空中的确存在着形影不离、互相绕转着的“星界鸳鸯——双星”。他在一生中发现了848对双星、三合星和聚星，并证实了维系着双星的是牛顿的万有引力理论，其运动则遵循着开普勒定律。其实，大约一个半世纪以前，有人已发现这种成对的恒星。只不过当时人们认为这些星之所以靠得那么近，是因为它们几乎恰好位于同一条视线方向上，而实际上，“双星”中两颗恒星相距很远。倘若情况真如此，那么，与较远的那颗星相比，较近的这一颗就应该显示出视差位移。赫歇尔对此做了大量观测，发现两颗星都未显示出视差位移的现象。根据它们的运动方式，赫歇尔认为“它们不是看上去黏在一起，而且实际上也的确靠得很近。”直至1793年，在大量观测的基础上，他确信成对的两颗恒星是在相互绕转。“双星”——他当时这样称呼它们，这个名称沿用至今。赫歇尔还非常仔细地观测了那些光度有变化的恒星，他是第一个系统地报道变星的

人。

赫歇尔的贡献几乎涉及天文学的所有领域。在太阳系中，除了天王星外，他还发现了4颗卫星：天卫三、木卫四、土卫一和土卫二。通过数十年如一日的1083次单调枯燥的恒星计数工作，他从60万颗恒星的测量中证明了银河系的存在，探知了它的形状、结构与大小，并用统计法首次确认了银河系为扁平状圆盘的假说。

此前，伽利略刚刚把望远镜指向夜空，就发现了很多用肉眼看不见的恒星。银河那白茫茫的光带中，原来充满了恒星。后来，赫歇尔把望远镜每改良一次，就能发现一大批更多、更暗的恒星。他通过对星空所做的系统观察发现，恒星在有些方向上多，有些方向上少。但他并不满足于这种定性的判断。1784年，赫歇尔决心要数一数天上的星星究竟有多少，并且想了解在不同的地方，星星的数目究竟是怎样分布的。要数清天上的星星，那可不是一件容易的事。耐心的赫歇尔首先把天空均匀分成几百个区域，然后数出每一个区域中用望远镜能看到的恒星。结果，赫歇尔发现，越靠近夜空中的那条乳白色的光带——银河，每单位面积上的恒星数目就越多；在银河的方向上达到最大值，而在与银河平面垂直的方向上，星星数目最少。

对于这种现象应该怎样解释呢？赫歇尔经过研究分析认为，恒星均匀地分布在形状如一个“透镜”或者一块“磨盘”那样的空间里，而我们的太阳系可能大约位于靠近中心的地方。而地球人朝着“透镜”直径方向看去，便可以看到一些较近的，因而较暗的星星，在外面是数目更多的更远、更暗的星。而大量十分遥远的星星由于亮度太暗、肉眼不可能一一分辨出来，只能看到白茫茫的光带，即银河。赫歇尔就是这样，用统计恒星数目的方法证实银河系为扁平状圆盘的假说。他开创了对银河系结构的研究，绘制了第一张银河截面图。尽管限于当时的条件，他的一些

结论并不完全正确，但无疑他是真正的“恒星天文学之父”，是开创银河系研究的先行者。

“在这个运动的宇宙里，为什么只有太阳一个是静止的呢？”赫歇尔就是抱着这样一个想法，开始研究太阳的空间运动状况的。远在公元8世纪初，中国唐代杰出的天文学家一行就把自己测量的恒星位置与汉代测量的位置相比较，发现存在变化。但他没有对此现象做出解释。1000多年后，英国天文学家哈雷用同样的方法（即参照古代记载的恒星位置），发现天狼星、大角星和毕宿五这3颗亮星有了明显的移动。为了研究上的方便，人们把恒星的运动分成两个分量：一个是视向速度，它在视线方向；一个是切向速度（即与视线垂直），表示为“自行”，即恒星每年在背景上位移的角度。通过对恒星运动的研究，1783年，赫歇尔发现了太阳的自行，他得到的太阳运动方向和现代测量数据相差不到 10° 。他指出太阳在银河系中也在运动着，即太阳率领着它的“子孙”，以每秒几十千米的巨大速度向着武仙座与天琴座毗邻的方向疾驰而去。

赫歇尔还在另一个方面扩展了人类的视野。1800年，他最早发现了太阳红外辐射。当时他用温度计测量太阳光谱的各个部分，发现在将温度计放在光谱红端外测温时，温度上升得最高，但那里却完全没有颜色。于是他得出结论：太阳光中包含着处于红光以外的不可见光线。红外天文学也由此发端起来……

赫歇尔对于天文望远镜的贡献更是无与伦比，他也是制造望远镜最多的天文学家。他共制作过400多架望远镜。赫歇尔利用全部业余时间制作望远镜，经过千锤百炼，他终于成为制造望远镜的一代宗师，他一生磨制的反射镜面达400多块。自古以来，人类对宇宙具有自然天生的敬畏和好奇心，这在赫歇尔身上有充分的体现。

从1773年起，他就亲自动手磨制镜头。这是一项极为枯燥又繁重的体力加智慧的工作。要把一块坚硬的铜盘磨成规定的极其光洁的凹面形，表面误差比头发丝还要细许多倍，中途还不能停顿，其难度可想而知。所以有时他要连续干上10多个小时，吃饭时只能由他的妹妹来喂他。开始时他连连失败了200多次，直到1774年他才尝到了胜利的快乐，制成了一架口径0.15米、长2.1米的反射望远镜，天王星的发现正是它的突出成果，他还看到了猎户座大星云和土星光环。在英王乔治三世的大力支持下，通过3年多的不懈努力，赫歇尔终于在1789年，51岁时，制造出了称雄世界多年的最大望远镜，它的镜筒直径达1.5米，差不多要3个人才能合围，镜筒长12.2米，竖起来有4层楼高，光是镜头就重2吨！这架像巨型大炮似的望远镜在使用的第一夜，就发现了土星的第一颗卫星——土卫二。

当时反射式望远镜的焦点多采用牛顿式，即在主焦点之前的光轴上，斜置一平面副镜，将焦点折射在镜筒上端的一侧。赫歇尔为了减少折射光的损失，将主镜略微偏置，使星光经主镜反射后，焦点不会汇聚在光轴上，而是斜到镜筒上端的一侧。这样，可以省去牛顿式的平面副镜，从而提高聚光的效率。因为这一光学系统是赫歇尔发明的，所以后世称之为“赫歇尔焦点”，而按照这种光学系统制成的望远镜称为“赫歇尔望远镜”。

从1781年到1782年的冬天，赫歇尔对星云基本构成的研究兴趣越发浓厚。他对星团和星云进行探测、研究，集20年观测成果，汇编成3部星云和星团表，共记载了2500个星云和星团，其中仅100多个天体是前人已知，并发现了一种新的天体——行星状星云。经反复仔细观测，他发现，他的高倍率望远镜能够辨别出几个星云团中的恒星个体。这一发现使他认为，星云之所以看过去是一片白茫茫的东西，是由于观测的

装备不够先进，只要有更精良的观测设，他们一定能分辨出其他星云里的恒星。这项发现促使赫歇尔在1784年和1785年提出，所有的星云都是由恒星组成的理论。其主要内容是：不必再用发光的奇妙流体来解释星云了，存在不能解析的星云是因为它们离我们太遥远的缘故；他提出存在着由众多恒星聚集为巨大天体的假设。后来，赫歇尔根据自己进一步观测到的行星状星云，确证了深空中确实存在弥漫状天体，它们是云气，而非原来认定的恒星。改正了其假设的错误。

赫歇尔当时通过把恒星密集的球状星团和疏散的星团做比较，发现这种结构形式还可以显示出引力作用的大小。他提出推论，即经过一段时间后，疏散星团必然会集中形成一个密集的星团和几个更紧密的星团。也就是说，恒星疏散的星团是由于星团处于早期发展阶段，而恒星密集的星团则属于星团的晚期阶段。因此，赫歇尔提出天体有“时间的变化”（即演化的），这一说法后来成为一个基本的科学概念。1785年，他发表了自己的天体演化理论。他指出，在广阔无垠的太空中，恒星最初是分散的，但随着引力的作用，渐渐聚集起来，形成了更加密集的星团。

1821年英国皇家天文学会成立时，他众望所归地成为首任会长，后来还被册封为爵士。赫歇尔成为专业天文学家时已经43岁，从未受过正规的高等教育，他的渊博学识、数理基础、冶炼技艺等全凭勤奋自学得到。赫歇尔具有强烈的求知欲，努力学习英文、意大利文、拉丁文，同时广泛阅读牛顿、莱布尼茨等科学家的自然哲学、数学、物理学著作，还接触了光学。他对天文知识有着浓厚的兴趣，对夜空中的许多星座都非常熟悉。

1822年，威廉·赫歇尔与世长辞。有趣的是，他84岁的寿命恰恰就是他所发现的天王星绕太阳公转一周的时间。

赫歇尔一家可称为天文世家，他的妹妹卡罗琳·赫歇尔（1750—1848）也是一位了不起的女性、杰出的天文学家。她终生未婚，与哥哥朝夕相处50年。威廉·赫歇尔的许多发现中也有她的一份功劳，她独自也有不少成就：发现了14个星云与8颗彗星，对星表做了修订，补充了561颗星。直到1848年，卡罗琳·赫歇尔以98岁高龄去世。赫歇尔的独子约翰·赫歇尔（1792—1871）也是著名的天文学家。他是英国皇家天文学会的创始人之一，发现的双星多达3347对，发现了525个星团星云，记下了南天的68 948颗恒星。他于1849年撰写的《天文学纲要》是对当时天文学的最好总结，对全世界都有深远的影响。

4.2.5 儒勒·凡尔纳的航天科幻小说《从地球到月球》和《环绕月球》

儒勒·凡尔纳（1828—1905），是19世纪法国著名的作家，被誉为“现代科学幻想小说之父”。他的航天科幻小说《从地球到月球》和《环绕月球》凭借着对当时科学与技术发展状况的把握，对未来的天才构想，用生动、准确的科学语言，描述了人类飞向太空、飞向月球，又返回地球的愿景。凡尔纳向19世纪的读者展示了一个“科学奇迹”成为现实的理想世界。而20世纪，他的一些科学幻想真的成了现实。例如，阿波罗登月。现在飞船上的返回舱，和凡尔纳在19世纪所设想的十分相似！



《从地球到月球》和《环绕月球》简介：

美国南北战争结束后，巴尔的摩城大炮俱乐部（这是大炮发明家的俱乐部）主席巴比康提议向月球发射一颗炮弹，建立地球与月球之间的联系。法国冒险家米歇尔·阿当获悉这一消息后，建议造一颗空心炮弹，他准备乘这颗炮弹到月球去探险。

巴比康、米歇尔·阿当和尼切尔船长克服了种种困难，终于在18××年12月1日乘这颗炮弹出发了。但是他们没有到达目的地，炮弹并没有在月球上着陆，却在离月球2800英里（约合4506千米）的地方绕月运行。

这三位冒险家的命运如何呢？据剑桥天文台的观测，只有两种可能，月球的引力征服了这颗炮弹，三位旅行家最后到达目的地；另一种

可能是炮弹被束缚在一个固定的轨道上，永远环绕月球运行。

《环绕月球》是《从地球到月球》的续集。巴比康、尼切尔和米歇尔·阿当乘坐炮弹到月球去探险途中，遇见一颗在太空游荡的火流星，它的引力使炮弹逸出轨道，无法抵达月球。三位旅行家对自己的危险处境置之度外，却仔细地观测了月球的面貌，并作了笔记。他们乘坐的炮弹因本身的速度太大，最后飞往与月球和地球引力相等的死寂点的另一边，向地球降落后，坠入太平洋。三位旅行家被一艘军舰救起，并受到美国人民的热烈欢迎。本书通过他们的奇特经历，描绘了星际空间变幻无穷的绚丽景象，从而使读者获得丰富的科学知识和广阔的想象空间。

凡尔纳小说中的月球炮弹与阿波罗登月情况对照表：

	《从地球到月球》	阿波罗登月
宇航员人数	3	3
航速	36 000英尺 / 秒（10 973米 / 秒）	35 533英尺 / 秒（10 830米 / 秒）
航时	97小时13分20秒、	103小时30分
降落地点	相差十几千米	
发射点	佛罗里达卡纳维拉尔角	

4.2.6 小结

在这一时期，天体力学诞生并取得了重要成就。使天文学从单纯描述天体的视位置和几何关系，进入到研究天体的相互作用，即从单纯的天体状态进入到研究天体运动原因。这是人类认识宇宙的一次重大飞跃。紧接着，由于分光术、测光术、光谱分析技术和照相术几乎同时用于天文学，天文望远镜也有重大发展，为研究天体的物理性质、化学组成等提供了条件，导致天体物理学诞生。它使对天体运动的认识又产生

一次飞跃：从只研究力学运动，到研究物理和化学运动。

天体力学和天体物理学的诞生构成人类认识天体的两次重大飞跃。天体力学诞生基于重大理论突破；天体物理学诞生则基于技术上的突破。

“理性工具”和“实体工具”在一定条件下，分别在人类认识自然和改造自然的过程中起主导作用，又构成不可分割、相互促进的两部分。

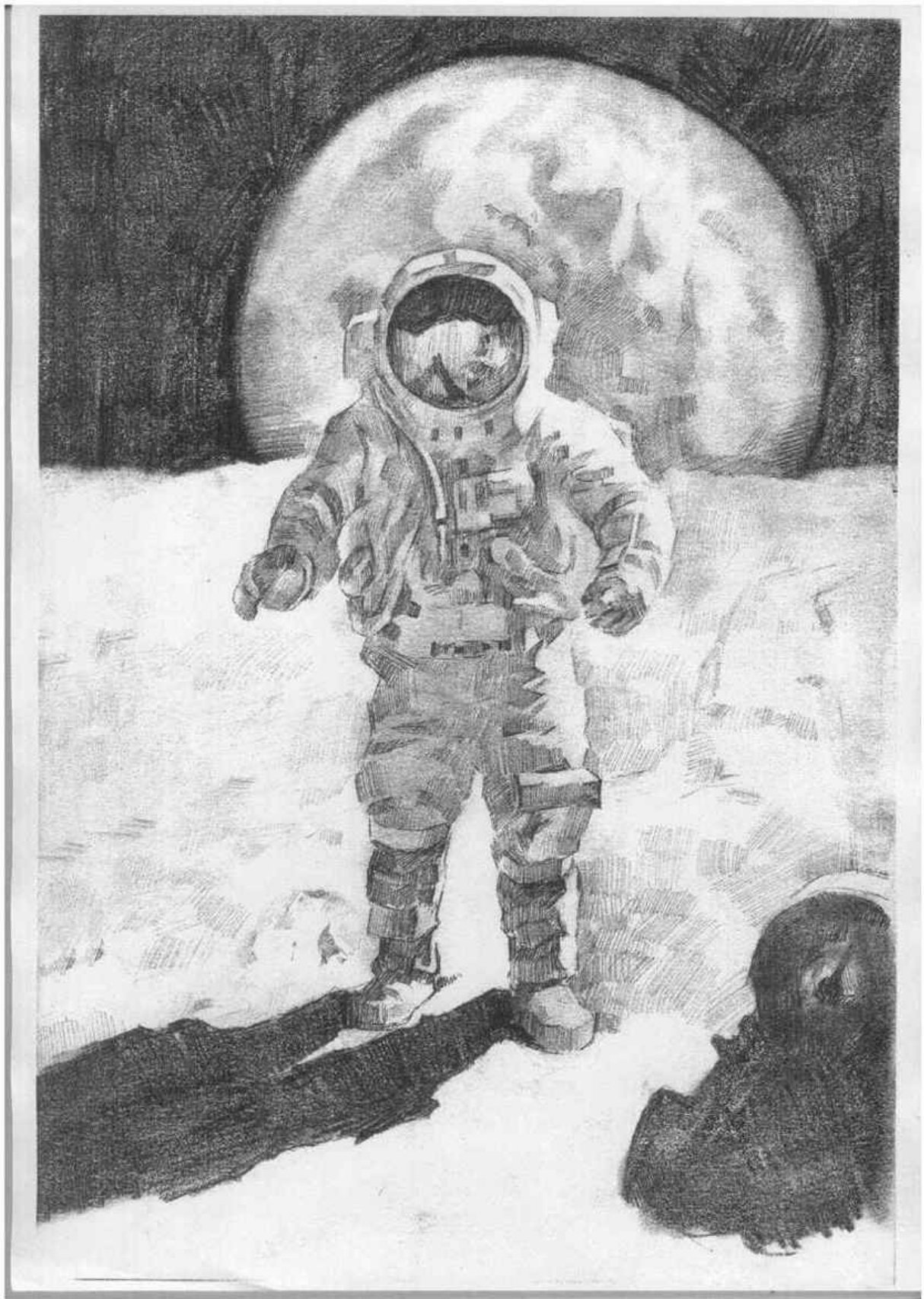
对恒星的科学认识、银河系概念的初步确立是人类对宇宙认识史上一个重大里程碑。人们的眼界从狭小的太阳系扩展到浩瀚的恒星世界，视野大为开阔。是进一步认识整个宇宙的一个阶梯。

从此，经典宇宙学建立起来，并为进一步的发展做好了准备。

参考文献

- [1] 温学诗, 吴鑫基. 观天巨眼: 天文望远镜的400年 [M]. 北京: 商务印书馆, 2008.
- [2] 邹海林, 徐建培. 科学技术史概论 [M]. 北京: 科学出版社, 2004.
- [3] 张邦固. 空间奥秘 [M]. 北京: 清华大学出版社, 2008.
- [4] 王鸿生. 世界科学技术史 [M]. 北京: 中国人民大学出版社, 2008.
- [5] 席泽宗. 世界著名科学家传记: 天文学家 [M]. 北京: 科学出版社, 1990.
- [6] 解启扬. 世界著名科学家传略 [M]. 北京: 金盾出版社, 2010.
- [7] [法] 弗拉马里翁. 大众天文学 [M]. 李珩, 译. 北京: 北京大学出版社, 2013.
- [8] [美] 温伯格. 宇宙学 [M]. 向守平, 译. 合肥: 中国科学技术大学出版社, 2013.
- [9] 钮卫星. 天文学史 [M]. 上海: 上海交通大学出版社, 2011.
- [10] 百度网. 儒勒·凡尔纳 [OL]. 2 2015-06-11.
<http://baike.baidu.com/view/25670.htm>

5 现代宇宙、天文学理论与宇宙实验观测证实



阿波罗登月

5.1 银河系之争及河外星系的确认

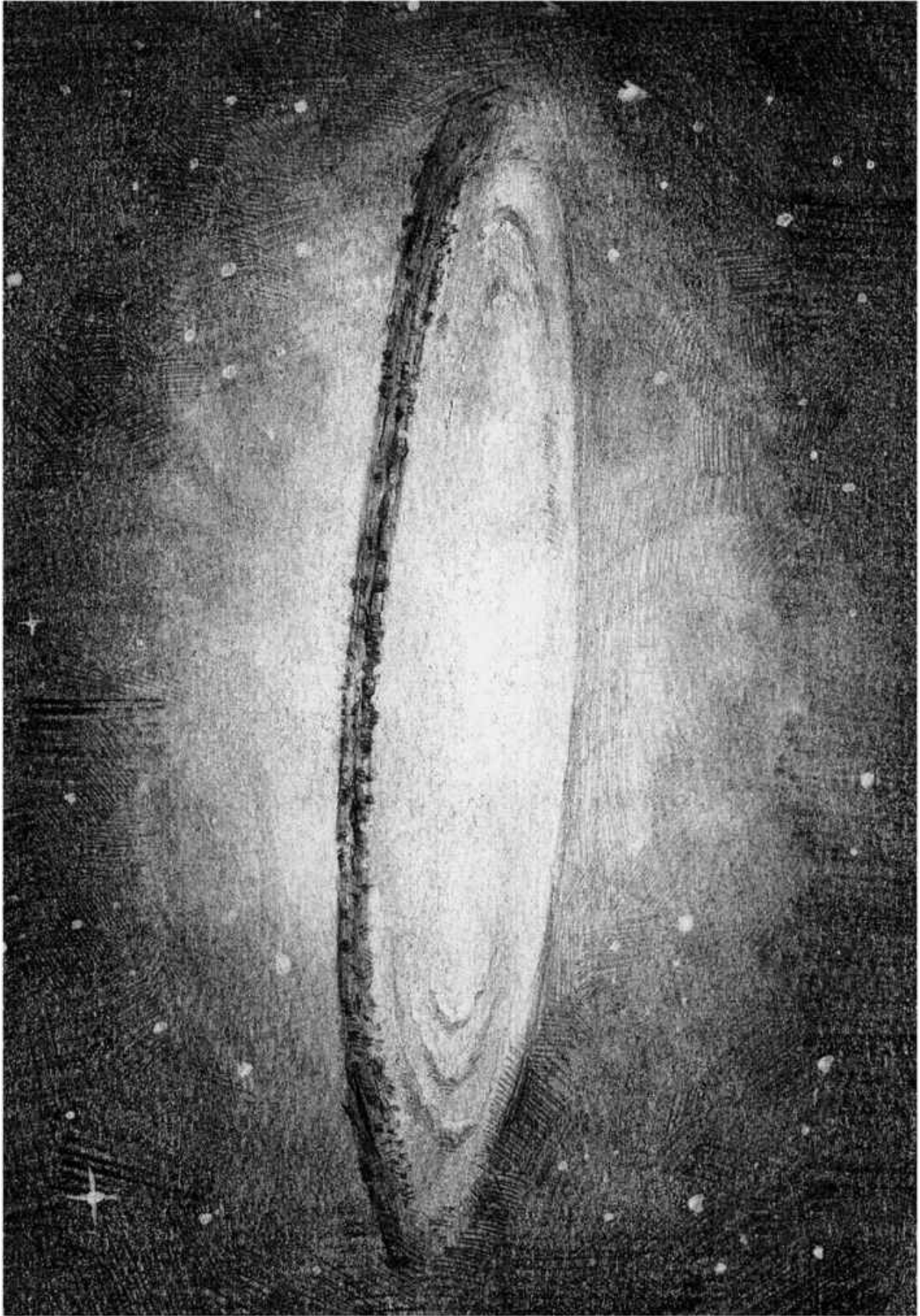
5.1.1 天体距离的观测技术和分析方法

从18世纪中叶到20世纪初，天文学者努力观察、探测星云和星团，以认清云雾状天体的本质和银河系的大小形状，确认了由气体和尘埃组成的气团和由恒星聚集成成的星团，但却无法分出是在银河系之内还是之外。

如何确定云雾状天体为河外星系？其必需的步骤是：

- （1）确定银河系大小和测定旋涡星云的距离；
- （2）由旋涡星云距离进一步定出它的大小；
- （3）最后确认它是否由大量恒星聚合而成。

当时，受限于观测技术和分析方法，人们对银河系构造的了解还相当模糊，自从赫歇尔之后，对银河系的研究便没有大的进步。当时对恒星距离的分析方法只有三角法（也叫“三角视差法”，是通过观测恒星的周年视差与地球观测者构成的三角形，从而得出恒星的距离）与恒星的自行测定（也叫“移动星团法”，根据恒星的运动速度来确定距离）。三角法只能测量距离小于300光年的星星。分光视差法与造父视差法后来陆续被发明，但能够反映恒星距离的只有自行。离得近的恒星有较大的自行，这一原则总体上必定是对的。“如果能够获得足够多的自行资料，那么在统计上可以应用这一原则，来给出一个达到更深空间的测量标杆。”



绘画：张京

1. 测定旋涡星云距离的探索

（1）三角视差法

1907年，瑞典天文学家波林（1869—1940）用三角视差法测得仙女座大星云的三角视差为 $0.171''$ ，相当于19光年。

（2）新星视亮度法

1911年，美国物理学家威里把仙女座大星云中一颗新星与英仙座新星比较，由此定出仙女座大星云的距离是1600光年。

（3）视向速度和自行法

1914年，美国天文学家斯莱弗（1875—1969）刊布了13个旋涡星云的视向速度。

1915年，美国天文学家柯蒂斯（1872—1942）测定了66个旋涡星云的自行。柯蒂斯认为从统计平均观点，天体视向速度与切向速度大致相同。求得旋涡星云自行平均值为 $0.033''$ ，再以星云视向速度平均值，求得星云平均距离为10 000光年。

2. 造父变星法求天体距离

1784年古德里克发现“造父一”亮度以5.37天为周期变化。

通常以周期1至50天光变的变星称为“造父变星”。

（1）造父变星的周光关系

20世纪初，美国女天文学家勒维特（1868—1921）用测光方法发现小麦哲伦云中许多变星。

1908年，她把周期长于1.2天的变星按亮度排列，结果发现周期也按大小排列。1912年，她发表周期为2至120天、视星等为12.5等至15.5等的变星资料，提出视星等和周期的对数存在正比关系，这就是周光关系。

（2）绝对星等与视星等的关系

1902年，卡普坦提出绝对星等概念。

对于视星等 m 和绝对星等 M ，现在通用如下公式：

若不计星际消光： $m - M = 5 \lg r - 5$ 。

若计及星际消光： $m - M = 5 \lg r - 5 + A(r)$ 。

其中， r 为天体距离（单位：秒差距）。

（3）周光关系零点的测定

丹麦天文学家赫茨普龙（1873—1967）指出勒维特在小麦哲伦云中发现的变星是造父变星。他进一步指出，在银河系中求一颗造父变星的距离 r 、视星等 m 和光变周期 P ，即可把勒维特周光关系图中的纵坐标由 m 改为 M 。这就是周光关系的零点问题。

1915年，沙普利用11个造父变星的自行和视向速度资料，求得造父变星统计视差，得到 $\lg P$ 与 M 的关系图。

20世纪40年代，德国科学家巴德（1893—1960）拍摄仙女座大星云

照片，核心部分也分解为恒星，证明旋涡星云是恒星系统。

5.1.2 银河系之争

在新的观测计算方法出现之后，宇宙学的发展进入了高速时期。首先引起学界争论的，便是银河系的结构与大小。1913年，天文学家赫兹普龙使用仅有的几颗银河造父变星的绝对星等，测定出了大、小麦哲伦星云的距离，大约是10千秒差距^[1]。尽管实际值约为此数的5倍，但却开创了一种新的测算方法，况且，以前天文学家从未测量过这样远的距离。

在此背景下，天文学家沙普利开始采用造父视差法探究球状星团的距离。在我们银河系内的球状星团，多数被发现在银河核心附近，并且在天球上的位置也大多数躺在银河核心周围的天空中。在1918年，沙普利便是利用这种强烈的不对称性推测星系的总体大小。他假设球状星团大致分布在银河核心的附近，经由球状星团的位置估计太阳与银河核心的距离。他得到的结果是距离约为30万光年，虽然他当时估计的距离有极大的偏误，但依然显示出星系的尺度大于早先的认知。他的偏误肇因是由于银河系内的尘埃减少了抵达地球的球状星团的光度，因而使距离显得更远。然而，沙普利估计的数值是在相同的数量级内，依然在现在可以接受的数值范围内。

沙普利的测量同时也指出太阳位于远离银河中心的位置上，反对早先从一般恒星的均匀分布所推导出来的结果。实际上，散布在银河盘面上的一般恒星经常会因为气体和尘埃的遮蔽而变暗，而球状星团分布在银河盘面之外，即使在更远的距离上仍然能被看见。

而1922年，卡普坦在分析已有的恒星资料后，提出了一个银河系的

模型，这个模型相较于赫歇尔的模型已经有了很大的改进，按照这个模型，银河系如同一个扁平的圆盘，中心便是一个恒星团，而太阳便在这个直径4000光年的星系的中央。这与沙普利的模型有较大不同，二人发生激烈争执，许多人怀疑沙普利的主张。

1927年，卡普坦的最后一个学生奥尔特证明了银河系的自转，他指出太阳及邻近恒星在以270千米/秒的速度，绕着3万光年外的银河系中心，做着周期2亿年的公转。

1930年，银河系的大小终于有了令人信服的解释。瑞士天文学家汤普勒在克里天文台研究银河星团时，“证明了星际空间不是像以前想象的那样完全透明，它们实际到处都有很稀薄的物质。”这些星际物质有如薄雾，能够吸收较远恒星的光线，使其看起来更远些。经过星际吸光效应的校正后，汤普勒得出太阳到银河系中心的距离是3万光年，同时银河系的直径也缩短到10万光年以下。

5.1.3 河外星系的确认

河外星系是否存在的问题，一直到20世纪20年代后期才得到解决。实际上，河外星系（如仙女座大星云）早就被观测到了，只是迫于当时的技术水平与测算方法，天文学家们并不知道这些星云状物质便是河外星系，只当是银河系内的某些星云。

1917年，美国天文学家柯蒂斯在旋涡星云中找到许多新星，他假定在这些星云中，新星的亮度极大时的绝对星等与银河系中的新星一样，由此估算出仙女座大星云的距离为1000万光年，后来减为50万光年。

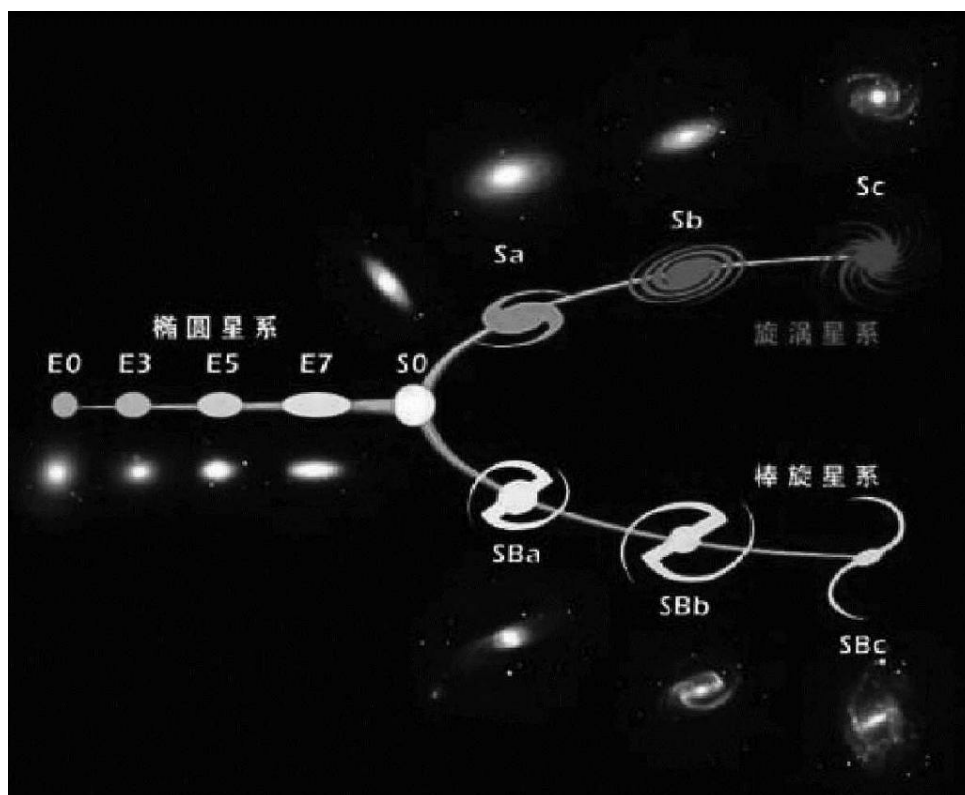
1923—1924年，哈勃通过威尔逊山天文台的100英寸（约合2.54

米)的反射望远镜,在仙女座大星云中找到并确认了一颗造父变星,根据这颗光变周期长达31天的造父变星,哈勃测算仙女座大星云到地球的距离,足有100万光年。即便是按照当时沙普利对银河系大小的估计

(沙普利估计银河系大小大约为30万光年),这个星云也远远位于银河系之外。之后,哈勃在仙女座大星云中发现了更多的造父变星,并在三角座M33与人马座星云中再次发现一些造父变星,他定出这三个星云的造父视差,估计出旋涡星云的距离约为285千秒差距,证明它们远在银河系之外。同时哈勃得出结论,仙女座大星云(此时应该称作“仙女星系了”)的直径只有银河系直径的1/10,体积只有银河系的1/1000。

但哈勃的同事范马宁在更早的时候便有了与哈勃的结果不相容的证据。他通过比较前后的底片,宣称发现了旋涡星云的自转。如果星云在短时间内能有察觉得到的自转,那么必定是小且近的,如果它们距离遥远,有着巨大的直径,可想而知,星云外围部分的自转速度将快到难以置信。这个发现与哈勃的结果格格不入,使得哈勃对自己的结果缺乏信心,直到1925年才经由他人之手发表。

1935年,哈勃通过观测彻底推翻了范马宁的结果,确立了仙女星系作为河外星系的地位。但哈勃对这个最大的河外旋涡星系的直径进行测量后,却只得到3万光年这一数字,远小于银河系直径。即使1930年汤普勒尔关于星际消光的理论被公布,两大星系的直径有所调整,但银河系仍然是最大的旋涡星系。天文学家对于任何满足人类虚荣心的理论都带着天生的反感。经过不断观测,1952年,侨居美国的德国天文学家巴德给出了新的周光关系,修正了仙女星系的数据。仙女座的距离较之前哈勃所观测的90万光年增加了一倍多,为230万光年,其直径也相应加倍,达16万光年左右。相比之下,银河系如同仙女星系的小妹妹一般。



在确认了第一个河外星系仙女大星系之后，哈勃又开始了对星系的分类研究。1926年，哈勃在论文中发表了他的星系分类法。他对主要的星系类型——椭圆、正常旋转和异常旋转——进行分析。椭圆星系按照其椭圆度进行排序；涡旋的和棒旋状的又被分为几个亚类；按涡旋结构缠绕的紧密度，以及星盘和星棒在星系的恒星分布中的重要性分别标以a、b和c。1936年，这个分类图最终以“音叉图”的形式呈现（如上图所示）。“哈勃将这个图看成是星系的演化过程，星系从图左侧的球状椭圆星系开始演化，最后进入旋转星系序列。这种思考后来被证明是不正确的（实际上椭圆星系要比旋涡星系要古老得多）。”但哈勃这种星系分类法却被宇宙学家们接受，不断得到完善，于20世纪50年代基本完成了分类工作。

5.2 现代宇宙学理论和观测证实的发展

5.2.1 宇宙膨胀说的提出

哈勃（1889.11—1953.9），美国天文学家，观测宇宙学的开创者。

哈勃对20世纪天文学做出许多贡献，被尊为一代宗师、“现代宇宙学之父”。其中最重大者有二：一是确认星系是与银河系相当的恒星系统，开创了星系天文学，建立了大尺度宇宙的新概念；二是发现了星系的红移-距离关系，促使现代宇宙学的诞生。他提出了哈勃定律（也叫红移定律）：河外星系的视向退行速度与距离成正比。根据多普勒效应，光的波长因为光源和观测者的相对运动而产生变化。在运动的光源前面，光被压缩，波长变得较短，频率变得较高，这便是蓝移；当运动在光源后面时，会产生相反的效应。波长变得较长，频率变得较低，这便是红移；光源的速度越高，所产生的效应越大。哈勃的发现和研究奠定了现代相对论宇宙学理论的基础。

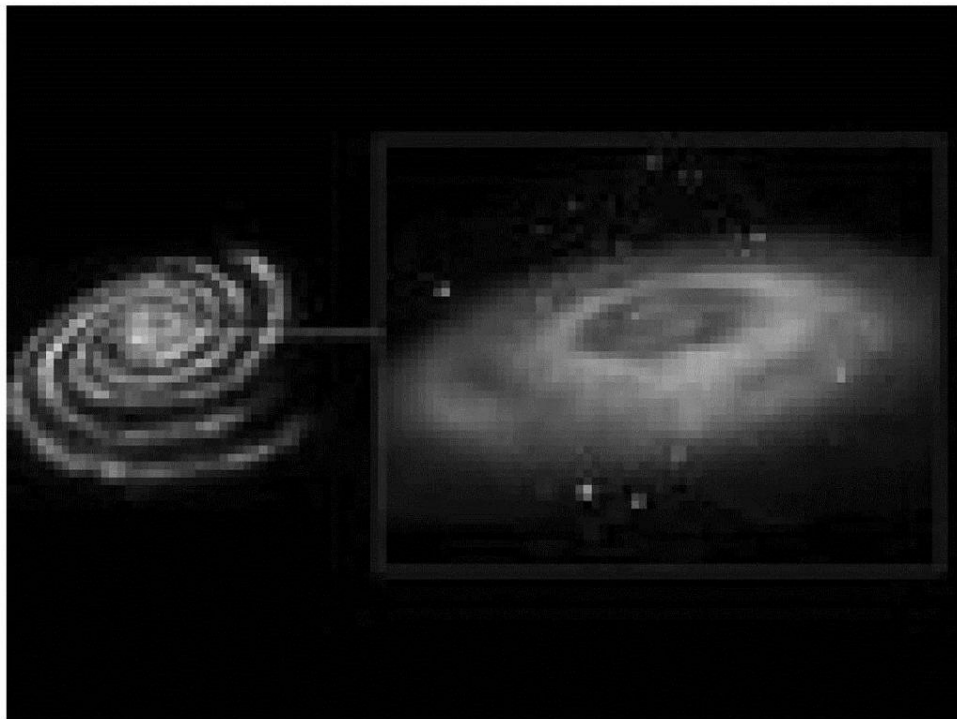


哈勃

1906年6月，17岁的哈勃高中毕业，获得芝加哥大学奖学金，前往芝加哥大学学习，在大学期间，他受天文学家海尔启发开始对天文学产生更大的兴趣。他在该校时即已获数学和天文学的校内学位。1910年，

21岁的哈勃在芝加哥大学毕业，获得奖学金，前往英国牛津大学学习法律，23岁获文学士学位。1913年在美国肯塔基州开业当律师。后来，他终于集中精力研究天文学，并返回芝加哥大学，25岁到叶凯士天文台攻读研究生，28岁获得博士学位。在该校设于威斯康星州的叶凯士天文台工作。在获得天文学哲学博士学位和从军两年以后，1919年退伍到威尔逊天文台（现属海尔天文台）专心研究河外星系并做出新发现。

1914年，他在叶凯士天文台开始研究星云的本质，提出有一些星云是银河系的气团。他发现亮的银河星云的视直径同使星云发光的恒星亮度有关。并推测另一些星云，特别是具有螺旋结构的，可能是更遥远的天体系统。



1919年，他用150厘米和254厘米望远镜照相观测旋涡星云。当时天文界正围绕“星云”是不是银河系的一部分这个问题展开激烈的讨论。

1923—1924年，哈勃用威尔逊天文台的254厘米反射望远镜拍摄了

仙女座大星云和M33的照片，把它们的边缘部分分解为恒星，在分析一批造父变星的亮度以后断定，这些造父变星和它们所在的星云距离我们远达几十万光年，远超过当时银河系的直径尺度，因而一定位于银河系外，即它们确实是银河系外巨大的天体系统——河外星系。1924年在美国天文学会一次学术会议上，正式公布了这一发现。这项发现使天文学家们关于“宇宙岛”的争论胜负立即分出，所有天文学家都意识到，多年来关于旋涡星云是近距天体还是银河系之外的宇宙岛的争论就此结束，从而揭开了探索大宇宙的新一页。1926年，他发表了对河外星系的形态分类法，后称哈勃分类。

20世纪初，斯里弗对旋涡星云光谱作过多年研究，发现谱线红移现象。在斯里弗观测的基础上，哈勃与助手赫马森合作，对遥远星系的距离与红移进行了大量测量工作，发现远方星系的谱线均有红移。只有极少数的星系谱线有蓝移，仅仅是银河系的三个邻居：M31（仙女座大星系）、M32（仙女座椭圆星系）、M33（三角星系）是在向与银河系接近的方向运动（蓝移），其他的星系都表现为离开银河系的红移，其中最大的红移值达到了0.12，根据波红（蓝）移的程度，可以计算出光源循着观测方向运动的速度。这意味着这个星系离我们远去的速度达到光速的12%——36 000km/s。而且距离越远的星系，红移越大，于是得出重要的结论：星系看起来都在远离我们而去，且距离越远，远离的速度越高。

1929年哈勃通过对已测得距离的20多个星系的统计分析，更进一步发现星系退行的速率与星系距离的比值是一常数。两者间存在着线性关系。这一关系被称为哈勃定律：河外星系的视向退行速度与距离成正比，每增加一百万秒差距（1Mpc，约为326万光年），星系的退行速度增加每秒500千米。哈勃第一次估算给出的这个比值，后来被称为哈勃

常数。哈勃定律的公式表达为

$$v=H \cdot d$$

式中， H 表示哈勃常数。

哈勃开创性地提出了哈勃定律，为我们推算宇宙年龄提供了新的方法。根据 $v=H \cdot d$ ，哈勃常数是一定值，而根据爱因斯坦相对论，星系的退行速度不可能达到或超过光速，则当我们假定星系退行速度达到最大——也就是光速时，星系的距离达到极限，而这也就是宇宙可能的最远的距离，也就是宇宙的年龄。因此，哈勃得到的宇宙的最大的距离是20亿光年，也就说明宇宙的年龄不会超过20亿年，然而这与当时根据同位素算得的地球的年龄46亿年相比，是有相当大的差距的。这里的差距并不是哈勃定律的问题，而是哈勃常数的问题，哈勃将哈勃常数定为 $500\text{km}/(\text{s} \cdot \text{Mpc})$ ，显然是有很大误差的。根据2010年NASA对最新观测数据的分析，哈勃常数的推荐值为 $70.8\text{km}/(\text{s} \cdot \text{Mpc})$ ，上下可允许误差为1.6，因此宇宙的年龄也被估算为138亿年。

“哈勃定律就展示了一幅宇宙整体退移也就是整体膨胀的图景：从宇宙中任何一点看，观察者四周的天体均在四处逃散，就像是一个正在胀大的气球，气球上的每两点之间的距离均在变大。”哈勃定律的提出为人类提供了一种新的宇宙观，鼓励着宇宙学家们提出新的宇宙模型，这被誉为20世纪最伟大的科学成就之一。

这一结论意义深远，因为一直以来，天文学家都认为宇宙是静止的。若认为红移是星系视向运动的多普勒效应造成的，则红移-距离关系表明，距离越远的星系正以越来越快的速度远离我们。运用广义相对论，人们通常把哈勃定律解释为宇宙膨胀的必然结果。哈勃定律的发现

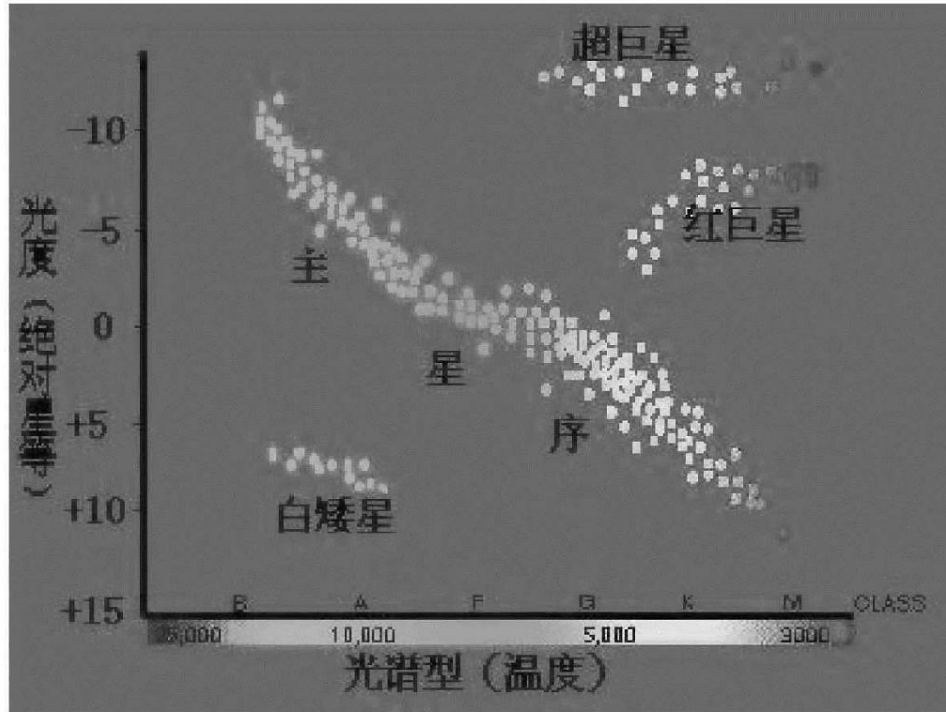
有力地推动了现代宇宙学的发展。

后来经过其他天文学家的理论研究之后，宇宙已按常数率膨胀了137亿年。

20世纪初，大部分天文学家都认为宇宙不会膨胀出银河系。但哈勃用当时最大的望远镜观察神秘的仙女座时，发现仙女座中的星云不是银河系的气体，而是一个完全独立的星系。在银河系之外存在极多其他的星系，宇宙比人类想象的要大许多。

5.2.2 赫罗图和恒星坍塌理论

20世纪，天体分光技术的运用、照相技术与测光技术的发明都在技术上极大地推动了天体物理学的发展，进一步促进了宇宙学研究。恒星的分光观测与光谱分类研究早在19世纪下半叶就已经有了阶段性成果，这直接推动了“赫罗图”的出现。



赫罗图是丹麦天文学家赫茨普龙及美国天文学家罗素分别于1911年和1913年各自独立提出的（见上图）。后来的研究发现，这张图是研究恒星演化的重要工具，因此把这样一张图以当时两位天文学家的名字来命名，称为“赫罗图”。赫罗图是恒星的光谱类型与光度之关系图，表示恒星的绝对星等与光谱型的分布，赫罗图的纵轴是光度与绝对星等，而横轴则是光谱类型及恒星的表面温度，从左向右递减。巨星分布于图片的右上方，再往上还有一些星，叫做“超巨星”。大多数星分布于从B型巨星到M型矮星的对角线型“主星序上”，太阳是其中的G型星，位于中间于偏右下的位置。赫罗图对天体物理学研究起到了极大的推动作用，“利用赫罗图，从恒星的光谱类型和视星等来确定恒星距离（分光视差）的逆方法，成了确定那些遥远的用三角学方法无法获得恒星距离的有力工具。”而这种依靠光谱分析法对天体进行研究的方法，很快也被宇宙学家用于对整个宇宙的研究。

之后的一种理论则是恒星坍塌理论，这个理论的最直接的来源便是

钱德拉塞卡极限。印度天文学家钱德拉塞卡计算得出白矮星的质量上限为1.44倍太阳质量，当恒星大于这个极限，将会坍缩成一种密度极大的状态，甚至是一个点。他把恒星起源归结于“引力的不稳定性”，这个奇怪的称谓到底是什么意思呢？

这是一个老的说法，它出现在20世纪20年代。其大意是，在均匀的气体系统中，如果出现了密度小扰动，它会像声波那样传播。因为引力存在，当此扰动的波长大于一个临界值时，它就可能呈指数变化，从而出现不稳定。

实际上，细致地分析会发现，这个奇怪的称谓根本不能解释恒星起源。

（1）从理论上讲，它是前后矛盾的。开始，它假设扰动是小的，在忽略了二次方以上的项之后，可以建立基本的波动方程。但后来，当扰动大到一定程度，比如说到了原密度的 $1/2$ 时，其二次方以上的项就不应该被忽略。也就是说，这时基本的波动方程已经不再成立。作为方程解的指数形式也就没有了基础。可是扰动怎么可能再继续大到破坏系统稳定的程度呢？

（2）实际上，它只是说扰动可能呈指数变化，可能变大，也可能变小。请注意，它只是说，可能，并且大和小的可能性同时并存。

（3）它没有指出，什么地方密度变大，什么地方密度变小。从其表达式来看，似乎是一起变化，大则一起大，小则一起小。

正由于它的随意性，便有了这样的表述：

宇宙大爆炸冷却下来以后，由于引力不稳定性，宇宙裂成许多大

块。它们就是将形成星系的气团。之后，又因为引力不稳定性，大块再裂成许多小块。这些就是恒星胎形成恒星的气团。进而，再次因为引力不稳定性，恒星胎形成恒星。

看看，引力不稳定性像不像童话里的精灵？它可以让星系胎之间的气体密度变小，同时让星系胎内气体密度变大一些；接着，再让恒星胎内一处气体密度陡然变大，而其他地方的气体密度变小（这就是所谓“坍缩”），以形成恒星。

引力从来都是起有序、稳定作用的。它怎么会不稳定？显然，这种“引力不稳定性”是不能从理论上解释恒星起源这个重大的科学问题的。这是“恒星坍塌理论”的硬伤。爱丁顿和爱因斯坦都宣布反对这个理论。

关于恒星起源的问题，中国的学者也开展了多方面的研究。其中较为符合观测事实的当推1994年由张邦固在《恒星起源运动学》（该书已在美国译成英文出版）一书中提出的理论。大爆炸以后，我们宇宙一直都浸泡在背景辐射之中。每一个粒子周围都有大约 10^{10} 个光子。所以，气体团应该处于热运动之中，粒子都有无规热运动速度。这部分速度与其位置无关。微观地看，粒子被与其位置有关的合引力加速后，它的与位置有关的速度越来越大。然而，它会与光子或者其他粒子发生碰撞。由于碰撞时，双方速度的大小和位置都是随机的，碰撞后，它们的速度也是随机的。也就是说，发生碰撞之后，粒子的速度便与它的位置无关了。

总之，在宇宙观系统里，在恒星胎中，粒子的速度应该分成两部分。一部分是热运动，它是无规则的，与位置无关的。在平衡态下，它服从麦克斯韦分布。另一部分是由合引力产生的，它是有规则的，指向

球心，依照牛顿第二定律增大，一直到发生碰撞。所以，它不断由零变大，与位置有关。

有了这种不同于宏观系统的统计力学思想，在玻尔兹曼微分积分方程基础上，便会得到一个新方程。

《恒星起源动力学》介绍了这个方程在几种初始条件下的求解过程。这里只给出结果。

1. 均匀气团

先从最简单的情况做起。这时的气团是一个均匀分布的气体球，处于平衡态。之后，其内部粒子之间的引力起作用，会出现本系统特有的局部运动，可能会离开平衡态。

在这种初始条件下，求解新方程，得到了下述结果：

(1) 临界半径： $C_r^{-1} = (3kT/4\pi m^2 n_0 G)^{1/2}$ 。这里，沿用了《恒星起源动力学》书中的符号。其中， k 是玻尔兹曼常量， T 是气团初始温度， π 是圆周率， m 是粒子质量。 n_0 是气团粒子数初始密度， G 是引力常量。

在解方程的过程中，临界半径自动地出现了。当气团的初始半径 r_0 大于它时，气团粒子就向中心汇聚，成为恒星胎。就是说，气团中心的密度随着时间变大而变大；而其他处的密度会逐渐变小。反过来，如果气团的初始半径小于它，那么气团中心密度会逐渐变小，气团会逐渐散掉。

从其表达式可以看到，它与气团温度的平方根成正比，与粒子数密度的平方根成反比。就是说，气团越热，要成为恒星胎的孤立气团的半

径要求越大。而密度越大，临界半径越小。

一般，温度大约是50开，密度大约是 10^{-22} 克 / 厘米³。这样，临界半径大约是8光年。银河系内恒星之间的平均距离大约是10光年。二者基本相符。

大家都知道，地球保持不住大气中的氢气。其原因是，与氢分子的热运动比较起来，地球的引力太弱。一般地说，一个引力系统维系其成员的条件是，在系统边缘处，成员的引力势能的绝对值大于其热运动的动能。统计力学告诉我们，一个氢分子热运动的动能是 $5kT/2$ 。而在由氢分子所组成的气团的边缘处，它的引力势能的绝对值是 $4Gm^2 n_0 \pi r_0^3 / 3$ 。于是，均匀气团不散失的条件是 $4Gm^2 n_0 \pi r_0^3 / 3 > 5kT/2$ 。这个条件与上面气团初始半径大于临界半径的条件大体上是一致的。这个条件告诉我们，孤立气团向中心汇聚的条件是，它的半径必须足够大，必须包含足够多的物质。显然，这个条件是完全符合物理实际的。

(2) 球心处粒子数密度

气团初始半径大于临界半径的条件时，均匀气团会在自身引力作用下收缩。其中心处密度按下式变化： $\rho(r=0, t) = \rho_0 / (1 - d_t^2 t^2)$ ，其中， t 是时间， ρ_0 是气团初始密度， d_t^{-1} 叫做恒星诞生时间。 $d_t = e^{-$

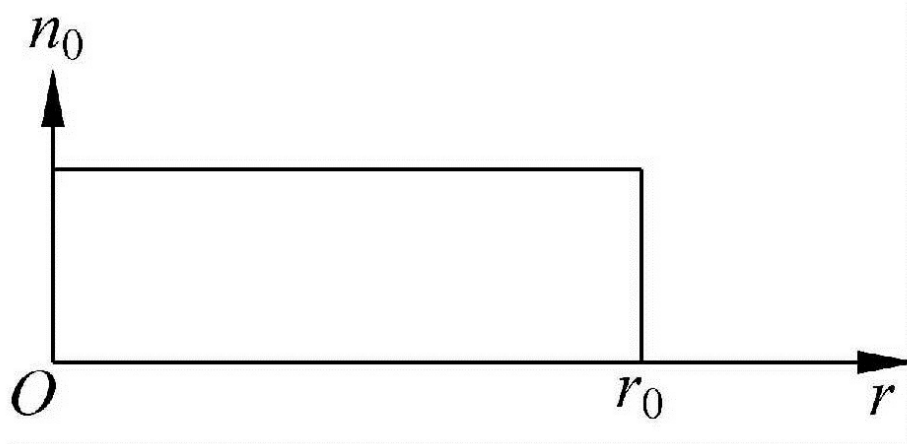
$1/24 (4\pi m n_0 G/3)^{1/2}$ ，其中， e 是自然对数的底。

显然，当时间达到诞生时间，气团中心处密度会达到生成恒星所必要的值。

诞生时间只与密度有关，对应于上面给出的密度，它大约是700万年。它也与天文观测基本相符。

2. 分层球对称

孤立系统是不与周围环境交换物质和能量，也不发生相互作用的系统。孤立气团也是这样。然而，孤立气团是出现在周围环境之中的。它的物质分布与环境不可能断得那样干脆。一过边缘，粒子便一个也没有了，如下图所示。所以，气团中心密度比较大，边缘比较小。这样的初始条件比较符合实际。这样，《恒星起源动力学》叙述了分层球对称平衡初始条件下求解新方程。其结果与均匀情况没有区别。



均匀分布

不过，这时的 n_0 是气团中心处的初始粒子数密度。

3. 旋转气团

一些恒星是旋转的，例如太阳。《恒星起源动力学》叙述了具有初始角速度 ω 的孤立气团的求解过程。

主要结果有：

- (1) 旋转气团收缩的必要条件包括了无旋气团的收缩条件——临

界半径。

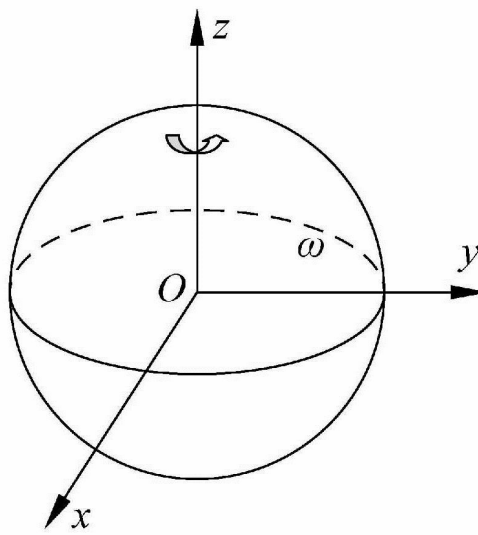
(2) 求解过程中，一个临界值

$$\omega_c = (2\pi\rho_0 G)^{1/2}$$

出现了。如果旋转气团的初始角速度大于这个临界值，则气团不会收缩。这个结果是有道理的。大家都知道，旋转运动会有离心力作用在粒子上。转得太快了，粒子就会被甩出去。系统就会消散，当然不会收缩。

(3) 在初始角速度小于上述临界值的条件下，气团会收缩，同时气团粒子还会向气团的赤道面聚集。赤道面是与气团角速度垂直、过球心的平面，就是下图的 xOy 面。初始角速度越大，“撒”到赤道面上的粒子越多。此特点显然与物理现实相符。

(4) 旋转恒星胎的诞生时间与初始角速度相关。如果初始角速度远小于临界值，那么，旋转恒星的诞生时间与无旋恒星的诞生时间差别不大。如果初始角速度大于临界值，那么，这时系统粒子的引力不敌离心力，恒星不会形成。可以认为，这时的诞生时间为无穷大。对于上述两种之间的情形，诞生时间会比无旋的长些。



赤道面

4. 膨胀气团

我们的宇宙在膨胀着。才从其中脱离出来的孤立气团也应该膨胀。膨胀着的孤立气团会在自身的引力作用下收缩吗？大量的恒星存在的事实给出了肯定的答案。《恒星起源动力学》求解了膨胀初始条件下的新方程。主要结果是：

(1) 在目前，膨胀对恒星形成的影响比引力小三四个数量级。也就是说，它基本上没有影响。

(2) 在宇宙膨胀比较剧烈的早期，膨胀给收缩条件加了一条：气团中心初始密度大于宇宙平均密度的2倍。

5. 有吸积核的孤立气团

《恒星起源动力学》介绍了有吸积核时方程的解。结果是，吸积核没有带来新条件。不过，这时的 n_0 是吸积核表面气体的初始粒子数密度。就是说，只要满足相应的条件，粒子就会向吸积核的表面（这个地方与其他气团有别）聚集。

《恒星起源动力学》的解释令人相信，满足相应条件的各种孤立气团都会成为恒星胎，在自身引力作用下收缩形成恒星。

现在问题是，在膨胀的宇宙中，各种孤立气团是怎样生成的？

对于这个问题，我们没有找到明确的答案。不过，根据观测的事实和已知的理论，是可以做一些推测和猜想的。

5.2.3 相对论与量子力学的发展

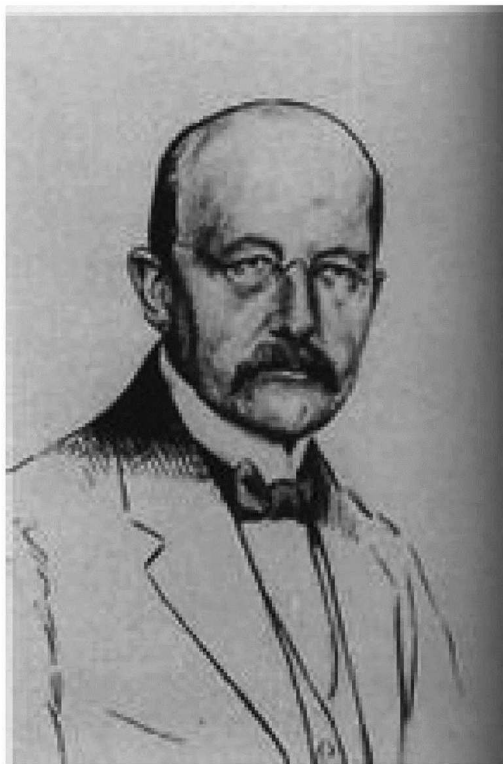
量子力学的创立实际要早于相对论的提出。1900年，德国科学家普

朗克在《关于正常光谱的能量分布定律的理论》一文中，正式提出了“能量子”的假说，认为能量不是无限可分的，并且能量并不是连续变化的，而是存在跳跃式变化。根据这个假说，他推导出了著名的“普朗克公式”。普朗克的能量子假说与牛顿经典物理学中信奉的连续的观念格格不入，当时学界大多数物理学家都持反对态度。“就连普朗克本人，在一个长的时期内，也对能量子假说认识不足，而犹豫徘徊甚至持怀疑态度，两次试图退回经典物理学。”但不管怎样，普朗克提出的量子概念是现代物理学中最重要的概念之一，他第一次把不连续的思想引入了物理学，使得物理学几乎所有领域都发生了根本性变革。

马克斯·普朗克（1858—1947），德国物理学家，量子论的奠基者，20世纪两位最重要的物理学家之一。

普朗克于1874年至1877年，在慕尼黑大学学习物理学和数学，1879年他转到柏林大学学习。1879年，他通过了博士论文，在论文中论述了热力学第二定律。1880年，他在慕尼黑大学担任物理讲师，1885年被基尔大学聘为理论物理特约教授。1900年，普朗克提出了一个重要的物理学常数——普朗克常数，以调和经典物理学理论研究热辐射规律时遇到的矛盾。

基于普朗克常数的假设，他推导出黑体辐射的普朗克公式，圆满地解释了实验现象。这个成就揭开量子力学的序幕，普朗克也因此获得1918年



马克斯·普朗克

的诺贝尔物理学奖。尽管在后来的时间里，普朗克一直试图将自己的理论纳入经典物理学的框架之下，但他仍被视为近代物理学的开拓者之一。1926年，普朗克成为英国皇家学会会员，同时还担任了柏林威廉皇家研究所所长。1947年10月逝世，终年89岁。

1900年，普朗克提的大胆假说在科学界一鸣惊人。这一假说认为辐射能（即光波能）不是一种连续不断的流的形式，而是由小微粒组成的。他把这种小微粒叫做“量子”。普朗克的假说与经典的光学说和电磁学说相对立，使物理学发生了一场革命，使人们对物质性和放射性有了更为深刻的了解。

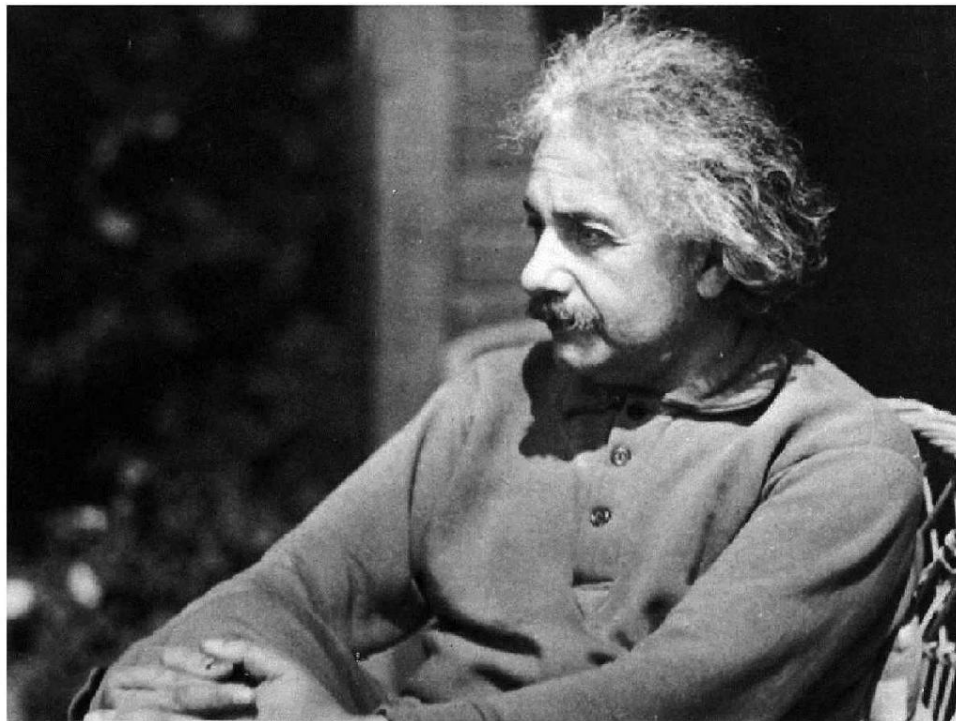
和其他科学家一样，普朗克对黑体辐射问题也很感兴趣，黑体辐射是描述给绝对黑体加热，来做电磁辐射的术语（绝对黑体是不反射任何光而完全吸收所遇见光的物体）。实验物理学家们甚至在普朗克着手研究这个问题之前就对这样的物体辐射做过认真的测量。普朗克取得的第一项成就是提出了一个用来正确描绘黑体辐射的相当复杂的代数公式。这个代数式完美地概述了实验数据，在今天理论物理学上仍常常使用。但是却有一个问题：公认的物理学定律预示存在着一个完全不同的公式。普朗克对这个问题沉思默想，终于提出了一个崭新的学说：辐射能只能以量子这个基本单位的整倍数形式辐射出来。根据普朗克学说，一个光量子的大小取决于光的频率（即颜色）且与一个物理量成正比。普朗克把这个物理量缩写为 h ，现在被称为“普朗克常数”。普朗克常数 h 在物理理论中有着重要的作用，现在被认为是两三个最基本的物理常数之一。它出现在原子结构学说、海森伯测不准原理、辐射学说和许多科学公式中。普朗克最初计算出来的常数数值与今天使用的相差2%。普朗克假说与当时流行的物理概念完全对立，但是他却利用这一假说在理论上准确地推导出正确的黑体辐射公式。普朗克假说具有彻底的革命性。

因此若不是他以顽固保守的物理学家而著称，他的假说无疑会被当作一种荒诞的思想而被弃之一边。

但是几年以后，他的假说被成功地应用到其他方面。1905年，爱因斯坦用这一概念解释了光电效应，1913年尼尔斯·玻尔在他的原子结构学说中也使用了这一概念。1918年普朗克获得诺贝尔奖。他的学说基本正确，而且在物理学理论方面具有重要的意义。量子力学的发展可能是20世纪中最重要的科学发展，甚至比爱因斯坦的相对论还要重要。

普朗克的另一个鲜为人知伟大的贡献是推导出玻尔兹曼常数 k 。他沿着玻尔兹曼的思路进行了更深入的研究后，得出了玻尔兹曼常数。为了向他一直尊崇的玻尔兹曼教授表示尊重，他建议将 k 命名为玻尔兹曼常数。

真正对量子概念起到巨大推动作用的是爱因斯坦，爱因斯坦在1905年发表的《关于光的产生与转换的一个启发性观点》一文中，明确提出了光量子的假说。他认为，即使在空中传播过程中，辐射也不是连续的，光是具有波粒二象性的物质。



阿尔伯特·爱因斯坦

阿尔伯特·爱因斯坦（1879—1955），德国物理学家，相对论的奠基者，20世纪最重要的两位物理学家之一。

赫兹于1887年发现光电效应，光照射到金属上，引起物质的电性质发生变化。这类光变致电的现象被人们统称为“光电效应”。光电效应分为光电子发射、光电导效应和光生伏特效应。前一种现象发生在物体表面，又称“外光电效应”；后两种现象发生在物体内部，称为“内光电效应”。1905年（26岁）3月，爱因斯坦发表量子论，提出光量子假说，成功地解释了光电效应。

金属表面在光辐照作用下发射电子的效应，发射出来的电子叫做光电子。光波长小于某一临界值时才能发射电子，此时的波长为极限波长，对应的光的频率叫做“极限频率”。临界值取决于金属材料，而发射电子的能量取决于光的波长而与光强度无关，这一点无法用光的波动性

解释。还有一点与光的波动性相矛盾，即光电效应的瞬时性，按波动性理论，如果入射光较弱，照射的时间要长一些，金属中的电子才能积累住足够的能量，飞出金属表面。可事实是，只要光的频率高于金属的极限频率，光的亮度无论强弱，光子的产生都几乎是瞬时的，不超过 10^{-9} 秒。正确的解释是，光必定是由与波长有关的严格规定的能量单位（即光子或光量子）所组成。光电效应里，电子的射出方向不是完全定向的，只是大部分都垂直于金属表面射出，与光照方向无关。光是高频震荡的正交电磁场，振幅很小，不会对电子射出方向产生影响。

早在16岁时，爱因斯坦就从书本上了解到光是以很快速度前进的电磁波。他非常想探讨与光波有关的所谓以太的问题。以太这个名词源于希腊，用以代表组成天上物体的基本元素。17世纪的笛卡儿和其后的克里斯蒂安·惠更斯首创并发展了以太学说，认为以太就是光波传播的媒介，它充满了包括真空在内的全部空间，并能渗透到物质中。与以太说不同，牛顿提出了光的微粒说。牛顿认为，发光体发射出的是以直线运动的微粒粒子流，粒子流冲击视网膜就引起视觉。18世纪牛顿的微粒说占了上风，但19世纪却是波动说占了绝对优势，以太的学说也大大发展。与此同时，电磁学得到了蓬勃发展，经过麦克斯韦、赫兹等人的努力，形成了成熟的电磁现象的动力学理论——电动力学，并从理论与实践上证明光就是一定频率范围内的电磁波，从而统一了光的波动理论与电磁理论。以太不仅是光波的载体，也成了电磁场的载体。直到19世纪末，人们企图寻找以太，然而从未在实验中发现以太，相反，迈克耳孙-莫雷实验却发现以太不大可能存在。

爱因斯坦似乎就是那个将构建崭新的物理学大厦的人。他认真研究了麦克斯韦电磁理论，特别是经过赫兹和洛伦兹发展和阐述的电动力学。爱因斯坦坚信电磁理论是完全正确的，但是有一个问题使他不安，

这就是绝对参照系“以太”的存在。他阅读了许多著作，发现所有试图证明以太存在的试验都是失败的。经过研究，爱因斯坦发现，除了作为绝对参照系和电磁场的荷载物外，以太在洛伦兹理论中已经没有实际意义。于是他想到：以太绝对参照系是必要的吗？电磁场一定要有荷载物吗？这时他开始怀疑以太存在的必要性。

爱因斯坦喜欢阅读哲学著作，并从哲学中吸收思想营养，他相信世界的统一性和逻辑的一致性。相对性原理已经在力学中被广泛证明，却在电动力学中却无法成立，对于物理学这两个理论体系在逻辑上的不一致，爱因斯坦提出了怀疑。他认为，相对论原理应该普遍成立，因此电磁理论对于各个惯性系应该具有同样的形式，但在这里出现了光速的问题。光速是不变的量还是可变的量，成为相对性原理是否普遍成立的首要问题。当时的物理学家一般都相信以太，也就是相信存在着绝对参照系，这是受到牛顿的绝对空间概念的影响。19世纪末，马赫在所著的《发展中的力学》中，批判了牛顿的绝对时空观，这给爱因斯坦留下了深刻的印象。

1905年5月的一天，爱因斯坦与一个朋友贝索讨论这个已探索了十年的问题，贝索按照马赫主义的观点阐述了自己的看法，两人讨论了很久。突然，爱因斯坦领悟到了什么，回到家经过反复思考，终于想明白了问题。第二天，他又来到贝索家，说：“谢谢你，我的问题解决了。”原来爱因斯坦想清楚了一件事：时间没有绝对的定义，时间与光信号的速度有一种不可分割的联系。他找到了开锁的钥匙，经过5个星期的努力工作，爱因斯坦把狭义相对论呈现在人们面前。

1905年6月30日，德国《物理学年鉴》接受了爱因斯坦的论文《论动体的电动力学》，在同年9月的该刊上发表。这篇关于狭义相对论的第一篇文章，包含了狭义相对论的基本思想和基本内容。不过文章刚发

表后，并没有立即引起很大的反响。后来普朗克注意到了他的文章，认为爱因斯坦的工作可以与哥白尼相媲美。正是由于普朗克的推动，相对论很快成为人们研究和讨论的课题，爱因斯坦也受到了学术界的注意。这一年因此被称为“爱因斯坦奇迹年”。

狭义相对论所根据的是两条原理：相对性原理和光速不变原理。爱因斯坦解决问题的出发点，是他坚信相对性原理。伽利略最早阐明过相对性原理的思想，但他没有对时间和空间给出过明确的定义。牛顿建立力学体系时也讲了相对性思想，但又定义了绝对空间、绝对时间和绝对运动，在这个问题上他是矛盾的。而爱因斯坦大大发展了相对性原理，在他看来，根本不存在绝对静止的空间，同样不存在绝对同一的时间，所有时间和空间都是和运动的物体联系在一起的。对于任何一个参照系和坐标系，都只有属于这个参照系和坐标系的空间和时间。

对于一切惯性系，运用该参照系的空间和时间所表达的物理规律，它们的形式都是相同的，这就是相对性原理，严格地说是狭义的相对性原理。在这篇文章中，爱因斯坦没有讨论将光速不变作为基本原理的根据，他提出光速不变是一个大胆的假设，是从电磁理论和相对性原理的要求而提出来的。这篇文章是爱因斯坦多年来思考以太与电动力学问题的结果，他从同时的相对性这一点作为突破口，建立了全新的时间和空间理论，并在新的时空理论基础上给动体的电动力学以完整的形式，以太不再是必要的，以太漂流是不存在的。

什么是同时性的相对性？不同地方的两个事件，我们何以知道它是同时发生的呢？一般来说，我们会通过信号来确认。为了得知异地事件的同时性，我们就得知道信号的传递速度，但如何测出这一速度呢？我们必须测出两地的空间距离以及信号传递所需的时间。空间距离的测量很简单，麻烦在于测量时间。我们必须假定两地各有一只已经对好了的

钟，从两个钟的读数可以知道信号传播的时间。但我们如何知道异地的钟对好了呢？答案是还需要一种信号。这个信号能否将钟对好？如果按照先前的思路，它又需要一种新信号，这样无穷后退，异地的同时性实际上无法确认。不过有一点是明确的，同时性必与一种信号相联系，否则我们说这两件事同时发生是无意义的。

光信号可能是用来对时钟最合适的信号，但光速非无限大，这样就产生一个新奇的结论，对于静止的观察者同时的两件事，对于运动的观察者就不是同时的。我们设想一个高速运行的列车，它的速度接近光速。列车通过站台时，甲站在站台上，有两道闪电在甲眼前闪过，一道在火车前端，一道在后端，并在火车两端及平台的相应部位留下痕迹。通过测量，得出的结论是，甲是同时看到两道闪电的。因此对甲来说，收到的两个光信号在同一时间间隔内传播同样的距离，并同时到达他所在位置，这两起事件必然在同一时间发生，它们是同时的。但对于在列车内部正中央的乙，情况则不同，因为乙与高速运行的列车一同运动，因此他会先截取向着他传播的前端信号，然后收到从后端传来的光信号。对乙来说，这两起事件是不同时的。也就是说，同时性不是绝对的，而取决于观察者的运动状态。这一结论否定了牛顿力学中引以为基础的绝对时间和绝对空间框架。

相对论认为，光速在所有惯性参考系中不变，它是物体运动的最大速度。由于相对论效应，运动物体的长度会变短，运动物体的时间膨胀。但由于日常生活中所遇到的问题，运动速度都是很低的（与光速相比），看不出相对论效应。

爱因斯坦在时空观的彻底变革的基础上建立了相对论力学，指出质量随着速度的增加而增加，当速度接近光速时，质量趋于无穷大。并且

给出了著名的质能关系式： $E=mc^2$ 。质能关系式对后来发展的原子能事业起到了指导作用。物质不灭定律，说的是物质的质量不灭；能量守恒定律，说的是物质的能量守恒。虽然这两条伟大的定律相继被人们发现了，但是人们以为这是两个风马牛不相关的定律，各自说明了不同的自然规律。甚至有人以为，物质不灭定律是一条化学定律，能量守恒定律是一条物理定律，它们分属于不同的科学范畴。

但爱因斯坦认为，物质的质量是惯性的量度，能量是运动的量度；能量与质量并不是彼此孤立的，而是互相联系、不可分割的。物体质量的改变，会使能量发生相应改变；而物体能量的改变，也会使质量发生相应改变。

爱因斯坦的质能关系公式，正确地解释了各种原子核反应：就拿原子量为4的氦来说，它的原子核是由2个质子和2个中子组成的。照理，其质量就等于2个质子和2个中子质量之和。实际上，这样的算术并不成立，氦核的质量比2个质子、2个中子质量之和少了0.0302u^[2]！这是为什么呢？因为当2个氘核（每个氘核都含有1个质子、1个中子）聚合成1个氦原子核时，释放出大量的原子能。生成1克氦原子时，大约放出 2.7×10^{12} 焦的原子能。正因为这样，氦原子核的质量减少了。

这个例子生动地说明：在2个氘原子核聚合成1个氦原子核时，似乎质量并不守恒，也就是氦原子核的质量并不等于2个氘核质量之和。然而，用质能关系公式计算，氦4原子核失去的质量，恰巧等于因反应时释放出原子能而减少的质量。

爱因斯坦从更新的高度，阐明了物质不灭定律和能量守恒定律的实质，指出了两条定律之间的密切关系，使人类对大自然的认识又深了一步。

1906年（27岁），爱因斯坦完成了固体比热的论文，这是关于固体的量子论的第一篇论文。1912年（33岁），提出“光化当量”定律。1913年，爱因斯坦应普朗克之邀，担任新成立的威廉皇帝物理研究所所长和柏林大学教授。

1914年（35岁），爱因斯坦接受德国科学界的邀请，迁居到柏林。8月，爆发了第一次世界大战后，他虽身居战争的发源地，生活在战争鼓吹者的包围之中，却坚决地表明了自己的反战态度。9月，爱因斯坦参与了发起反战团体“新祖国同盟”。

与此同时，爱因斯坦继续探索相对论的有关问题，在一篇文章中第一次提到了等效原理。此后，爱因斯坦关于等效原理的思想又不断发展。他以惯性质量和引力质量成正比的自然规律作为等效原理的根据，提出在无限小的体积中均匀的引力场完全可以代替加速运动的参照系。他还提出了封闭箱的说法：在一封闭箱中的观察者，不管用什么方法也无法确定他究竟是静止于一个引力场中，还是处在没有引力场却在作加速运动的空间中，这是解释等效原理最常用的说法。而惯性质量与引力质量相等是等效原理的一个自然的推论。

1915年（36岁）11月，爱因斯坦先后向普鲁士科学院提交了4篇论文。在这4篇论文中，他提出了新的看法，给出了《广义相对论》引力场方程的完整形式，并且成功地解释了水星近日点运动。至此，广义相对论的基本问题都解决了，广义相对论诞生了。1916年，爱因斯坦完成了长篇论文《广义相对论的基础》。在这篇文章中，爱因斯坦首先将以前适用于惯性系的相对论称为狭义相对论，将只对于惯性系物理规律同样成立的原理称为狭义相对性原理，并进一步表述了广义相对性原理：物理学的定律必须对于无论哪种方式运动着的参照系都成立。

爱因斯坦的广义相对论认为，由于有物质的存在，空间和时间会发生弯曲，而引力场实际上是一个弯曲的时空。爱因斯坦用太阳引力使空间弯曲的理论，很好地解释了水星近日点进动中一直无法解释的43秒。广义相对论的第二大预言是引力红移，即在强引力场中光谱向红端移动，20世纪20年代，天文学家在天文观测中证实了这一点。广义相对论的第三大预言是引力场使光线偏转。最靠近地球的大引力场是太阳引力场，爱因斯坦预言，遥远的星光如果掠过太阳表面将会发生1.7秒的偏转。1919年，在英国天文学家爱丁顿的鼓动下，英国派出了两支远征队分赴两地观察日全食。经过认真的研究，得出的最后结论是：星光在太阳附近的确发生了1.7秒的偏转。英国皇家学会和皇家天文学会正式宣读了观测报告，确认广义相对论的结论是正确的。会上，著名物理学家、皇家学会会长汤姆逊说：“这是自从牛顿时代以来所取得的关于万有引力理论的最重大的成果”，“爱因斯坦的相对论是人类思想最伟大的成果之一”。爱因斯坦在1916年写了一本通俗介绍相对论的书《狭义与广义相对论浅说》，到1922年已经再版了40次，还被译成了十几种文字，广为流传。

1955年（76岁）4月18日午夜，爱因斯坦逝世于普林斯顿。遵照他的遗嘱，他死后并没有举行任何丧礼，也不筑坟墓和纪念碑，骨灰撒在了永远保密的地方。

在此之后，德布罗意提出了物质波假说，波尔提出了关于几率波的假说，海森伯创立了矩阵力学，薛定谔创立了波动力学。在短短20多年内，量子力学便从初生到蓬勃发展再到形成完整的理论体系，成为现代物理学的理论基石，也促使宇宙学开创了从小到极致的粒子研究整个宇宙的方法。

相对论的意义：

狭义相对论和广义相对论建立以来，已经过去了很长时间，它经受了实践和历史的考验，是人们大都承认的真理。相对论对于现代物理学的发展和现代人类思想认识的发展都有巨大的影响。相对论从逻辑思想上统一了经典物理学，使经典物理学成为一个完美的科学体系。

狭义相对论在狭义相对性原理的基础上统一了牛顿力学和麦克斯韦电动力学两个体系，指出它们都服从狭义相对性原理，都是对洛伦兹变换协变的，牛顿力学只不过是物体在低速运动下很好的近似规律。广义相对论又在广义协变的基础上，通过等效原理，建立了局域惯性系与普遍参照系之间的关系，得到了所有物理规律的广义协变形式，并建立了广义协变的引力理论，而牛顿引力理论只是它的一级近似。这就从根本上解决了以前物理学只限于惯性系的问题。从逻辑上得到了合理的安排。相对论严格地考察了时间、空间、物质和运动这些物理学的基本概念，给出了科学而系统的时空观和物质观，从而使物理学在逻辑上成为完美的科学体系。

对于爱因斯坦引入的这些全新的概念，当时地球上大部分物理学家，其中包括相对论变换关系的奠基人洛伦兹，都觉得难以接受。甚至有人说“当时全世界只有两个半人懂相对论”。旧的思想方法的障碍，使这一新的物理理论直到一代人之后才为广大物理学家所熟悉，就连瑞典皇家科学院，1922年把诺贝尔物理学奖授予爱因斯坦时，也只是说“由于他对理论物理学的贡献，更由于他发现了光电效应的定律”。对爱因斯坦的诺贝尔物理学奖颁奖词中竟然对于爱因斯坦的相对论只字未提。应当承认，相对论没有获诺贝尔奖，一个重要原因就是还缺乏大量事实验证。

爱因斯坦的理论直接否定了经典宇宙学中“以太”假说，打破了时空的局限性与分割性，将空间物体的质量、速度与时间、空间相互联系起

来，并提出了时空弯曲假设，在引力场中抛弃了惯性系。但是，爱因斯坦错误地提出了“宇宙常数”作为静态宇宙模型的基础，认为“宇宙常数”所导致的微弱斥力，会与万有引力相抗衡，这样宇宙便消除了膨胀或收缩的全局性动态变化，而归于永恒状态。后来，在参观了威尔逊山天文台，与哈勃、勒梅特等人推心置腹地讨论各自观点后，爱因斯坦放弃了自己的学说，并承认引入“宇宙常数”是一生中最大的失误。然而，这并不妨碍狭义与广义相对论基本是正确的，之后宇宙学家的几乎所有理论都是基于相对论而提出的。

5.2.4 宇宙大爆炸理论

哈勃向世界证明了宇宙正如一个气球一般不断胀大，引起科学界的关注。同时，也有天文学家想到，宇宙既然是不断膨胀的，“这种情况也就意味着，过去的宇宙比今天的宇宙占有较小的空间尺度。因此，如果不断地以时间回溯，越早期的宇宙就会越小；那么总会有足够早的某个时刻，宇宙是处在它的最小的状态。”

比利时最著名的天文学家，也是一位天主教士——勒梅特（1894—1966），开创了研究宇宙起源的先河。1932年，他提出宇宙有一个起始之点，这个点正好与教会正苦苦追寻的上帝创世之点契合。他认为，上帝创世时创造了一个“原始原子”，而后这个“原始原子”不断胀大，如同橡果长成参天的橡树一般。而这个模型，也与爱因斯坦方程式中暗藏的宇宙膨胀趋势相符合。



根据哈勃定律进行时间回溯，勒梅特推算出宇宙的年龄大约是150亿年。勒梅特的“原始原子”假说如果得到证实，那么毫无疑问，“关于宇宙无限和永恒的观念是错误的”。滑入了形而上学的上帝创世论。

1. 大爆炸宇宙模型

1948年，俄裔美籍物理学家伽莫夫（1904—1968）与他的博士生阿尔夫提出了热大爆炸宇宙模型。他们认为宇宙在膨胀的初期存在过一个高温、高密的“原始火球”，在这个特殊状态中，同时存在着质子、中子、正负电子与中微子，各种粒子以极其致密的形式处于平衡状态。随着宇宙膨胀，温度降低，平衡过程被破坏。一部分中子因 β 衰变成为质子和电子，质子由于俘获中子成为重质子。这样，由于反复发生质子俘获与 β 衰变，更重的元素便由此产生。

热大爆炸宇宙模型后来成为科学界主流宇宙学基础理论。

2. 大爆炸理论的反对者

同年，剑桥大学教授邦迪、霍伊尔与戈尔德等人提出了“稳态宇宙模型”，这种模型虽然称作稳态，却并不排斥宇宙膨胀说。霍伊尔等人提出，宇宙在空间上均匀各向同性，在时间上恒稳不变；认为由于膨胀而在各星系间产生的缝隙中，不断会有新的物质（恒星）产生，填补膨胀所遗留下来的空间。这种学说克服了当时由于哈勃常数不精确，导致的宇宙年龄小于地球年龄的问题，以及光度的佯谬，成为与“大爆炸”理论相抗衡的学说。

为了使自己的学说更有说服力，霍伊尔与他的同事探究了恒星的演化过程。他们研究发现，当空间中的氢原子由于引力逐渐聚合到一起，会形成越来越大的球体。受引力作用，氢原子间的密度也不断随着增

大，内部压力越来越大。当压力达到足够高的程度，氢原子将发生聚变反应，形成氦元素。在核聚变的过程中，巨大而持续的能量被释放出来，恒星发出了强烈的光亮，这便解决了恒星为何发光的古老难题。

霍伊尔等人还指出，恒星在燃尽了氢元素之后，体内大多数都是稍重的氦元素，由于缺少氢元素发生核聚变，氦元素将以聚合得更加紧密。于是，氦元素之间由于巨大的压力发生新的核聚变，形成更重的元素。新的元素由于压力又会产生新的更重的元素。而氦元素合成新的元素之前，并非所有的氢元素都已经消失。每一次合成新的元素，都会有原有的元素剩余下来。这就意味着，只要恒星的质量足够大，在恒星的晚年便可以产生所有的重元素。霍伊尔大胆地预言，超新星爆发时极高的温度，能够引发不同寻常的核聚变，产生更多的重元素。后来的发现表明，超新星上存在着几乎所有种类的重元素。

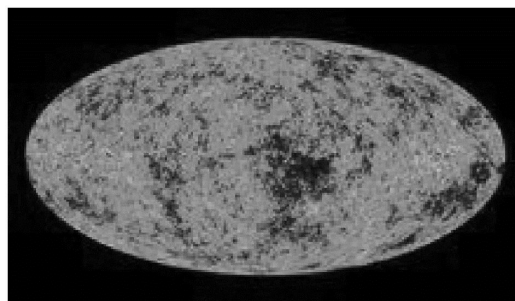
尽管“稳恒态宇宙模型”违背了能量守恒定律，没有说明物质产生的具体途径和机制，也无法解释背景辐射与氦元素丰度问题，但是霍伊尔等人对恒星演化过程的研究却是值得称道的。

3. 大爆炸理论最重要的证据

在研究宇宙早期状态时，大爆炸理论提出者阿尔夫发现，大爆炸早期的甚高温状态下，宇宙呈现的是辐射态而不是物质态。根据大爆炸理论，大爆炸1秒以后，宇宙处于高温、高热（大约100亿摄氏度）的粒子“羹汤”状态，这时整个宇宙处于均匀的热平衡状态。随着宇宙的膨胀和降温，其中的一些粒子逐次与其余部分粒子脱耦（由强关联关系变为弱关联关系）。此时产生的核反应使中子和质子聚合在一起形成氦核，余下的核子（没有聚合的质子）自然就形成了氢核。精确的理论计算表明，当时应有23.6%的物质质量聚合成了氦核。英国皇家格林尼治天文

台对众多星系中原始星云的发射光谱进行观测的结果表明，宇宙中氦的实际丰度为23.5%。这一结果与大爆炸的理论预言极为相符，成为大爆炸理论最重要的证据之一。而稳恒宇宙理论却无法对此进行解释。进一步研究表明，“这些早期炽热阶段的冷却遗迹就是我们今天所呈现的宇宙。”经过计算，他们推算出热背景辐射应该在5开左右。

发现背景辐射的不是正规的宇宙学家，而是贝尔电话实验室的两位工程师——彭齐亚斯和威尔逊。1964年，普林斯顿大学的罗伯特·迪克与他的同事们制造了一个精密的探测器，踌躇满志地准备开始寻找阿尔夫



所言的背景辐射，但却接到一个陌生的电话。这个电话来自贝尔电话实验室。当时的彭齐亚斯与威尔逊正在校正为测试卫星通信而设计的号角式反射天线。他们以极大的耐心追踪和去除了各种干扰源，“甚至清除了巨大牛角形天线中的鸽子粪便”，却发现干扰依旧如故。他们测定了那背景辐射的温度，得到的答案是3.5开。1989年，美国宇航局发射了宇宙背景探测者卫星（cosmic background explorer, COBE）。1990年，COBE的“远红外绝对分光光度计”小组宣布，宇宙微波背景辐射非常精确地等于绝对温度为2.725开的黑体辐射，上下误差为0.001开。这个结果被许多科学家看作是大爆炸学说无可争辩的观测事实。宇宙背景辐射——大爆炸的遗迹——就这样被找到了。彭齐亚斯与威尔逊也因此获得了1978年度的诺贝尔物理学奖。

宇宙背景辐射的发现为10多年来一直被讨论的“假说”提供了一项强有力的证据，包括在大爆炸理论提出前便被论证了的哈勃宇宙膨胀理论。再加上氦元素丰度（大爆炸理论推测宇宙中氢元素丰度大约是

75%，氢元素丰度为24%，而现在对宇宙的观测证实了宇宙大爆炸理论的推测，这个已经成为证实宇宙大爆炸理论的一个有力的证据）的证实，宇宙大爆炸理论得到了最重要的三个证据。这也说明了大爆炸理论的包容性与吸引力，尽管未被完全证实，却几乎已经成为一个真理。“对于许多宇宙学家来说，大爆炸理论今天成了天体物理学的聚合力量，它使天体物理学与粒子物理学相关联，也使整个天文学成为一个统一的整体。”大爆炸理论之于今日宇宙学，其意义之深远，再怎么估计都不会过分。

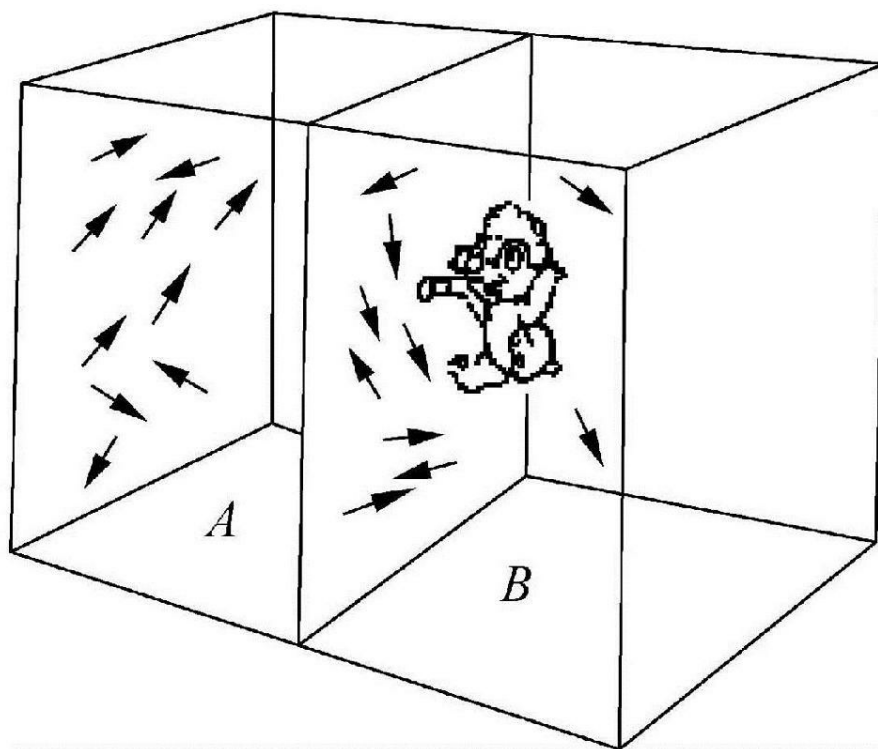
到了20世纪80年代，美国宇宙学家古斯提出了大爆炸模型的一种补充理论——暴胀宇宙模型。在这个模型中，在大爆炸后的10秒至30秒内，宇宙的尺度按指数攀胀，其间温度急剧下降后回升，视界距离疾增，物质向现有粒子形式转化。这个理论中以奇点代替了“原始火球”，解决了大爆炸模型无法解决的视界问题、平直性问题与磁单极子问题。但暴胀模型的尚不成熟，目前尚无完全成功的理论模型。暴胀宇宙模型实际是大爆炸说的一种改良。

还有学者提出“奇点起源说”，又可以叫“无中生有”宇宙观，大意如下：

目前观测到的宇宙大爆炸源于一个四维时空奇点。它没有大小，没有过去，没有物质，没有一切。突然，它变大了，有了时间，有了空间，有了能量、质量，有了一切。宇宙大爆炸开始了。

不过，这种理论并不科学：①它违反了普适的能量守恒定律、质量守恒定律等科学规律；②这个奇点没有来源，没有“因”。所以，它违反了作为科学基础的因果律。

大爆炸理论再次火热起来的同时，也引发了许多相关问题的激烈讨论，宇宙“热寂说”便是其中一个。热寂说博大精深，对热寂说的研究探讨也成为近代科学史中最令人费解的谜题。热寂说的提出者，是热力学第二定律的提出者，开尔文与克劳休斯。



麦克斯韦妖

热力学第二定律认为，不可能把热从低温物体传到高温物体而不产生其他影响；不可能从单一热源取热使之完全转换为有用的功而不产生其他影响；不可逆热力过程中熵的微增量总是大于零。进而推断出，在一切自然现象中，熵（物理学上指热能除以温度所得的商，标志热量转化为功的程度）的总值永远只能增加而不能减少。于是到处不断进行的变化过程，可以用下面的定律简短地表述：宇宙的熵趋于极大。宇宙越是接近于这个熵是极大的极限状态，那就任何进一步的变化都不会发生。

了，这时宇宙就会进入一个死寂的永恒状态。

热寂说提出后不久，便受到麦克斯韦的诘难，麦克斯韦设计了一个假想的存在物——“麦克斯韦妖”。麦克斯韦妖有极高的智能，可以追踪每个分子的行踪，并能辨别出它们各自的速度。这个设计方案如下：“我们知道，在一个温度均匀的充满空气的容器里的分子，其运动速度绝对不均匀。然而任意选取的任何大量分子的平均速度几乎是完全均匀的。现在让我们假定把这样一个容器分为两部分，A和B，在分界上有一个小孔。再设想一个能见到单个分子的存在物，打开或关闭那个小孔，使得只有快分子从A跑向B，而慢分子从B跑向A。这样，它就不消耗功的情况下，B的温度提高，A的温度降低，这与热力学第二定律发生了矛盾。”

热寂说不仅在科学上受到科学家的反对，恩格斯也曾反对过这一学说，使得这一学说的争议再次扩展到哲学上。但是100多年来，尽管众多科学家都曾提出过各式各样的反对意见，热寂说却始终未伤及根本。麦克斯韦虽然提出了“麦克斯韦妖”这一假想的生物，却并未设计实验证明这一假说的真实性。后世虽有许多反对的声音，实际都未能切中要害。直到20世纪60、70年代大爆炸说为众多天文学家接受以后，热寂说才受到了前所未有的挑战。

彭齐亚斯与威尔逊在发现宇宙背景辐射时，发现各个方向的噪声都是一样的，这便说明背景辐射温度在所有的方向上都是相同的。更进一步，这便说明宇宙的初始状态是均匀的、平衡的，符合热力学第二定律所说的热平衡状态。大爆炸之后，宇宙才逐渐偏离热平衡态。

热寂说在此时受到的挑战在于，在一个膨胀的宇宙中，是否能够存在热平衡呢？

以下便是一种根据大爆炸理论对热寂说提出的反对：宇宙中存在着两类基本物质，一种是辐射，一种是粒子。辐射的温度与粒子的温度不一样。那么按照经典热力学，经过一段时间后，二者温度必然完全相同。这是静态空间中的结论。然而在膨胀的宇宙中，结论却完全不同。在膨胀的过程中，辐射的温度降低较慢，粒子温度降低较快，会造成辐射温度高于粒子温度，从而产生温差。这与经典力学的结论是相反的。尽管这个温差会因为辐射与粒子的碰撞而消失，最终达到热平衡，但是由于辐射与粒子相撞的可能性较低，因而达到热平衡所需的时间远比宇宙膨胀所需的时间要长，因而辐射和粒子之间永远不可能达到热平衡。此时系统的熵尽管在不断增加（即热量不断丧失，符合热力学第二定律），但却离平衡态越来越远。

在一个体系中，如果同时存在着正热容物体和负热容物体，那么这个体系就具有极大的不稳定性。稍有扰动，平衡就会彻底遭到破坏而产生温差。只要有自引力体系存在，原则上就不存在稳定的热平衡，而宇宙间的天体或天体系统大多数正是这种自引力系统。尽管自引力系统中熵是增加的，但由于没有热平衡，因而熵的增加是无止境的，永远都没有极大值。

热力平衡是热力学第二定律的基石，在引力起作用的体系中，热力学意义上的热平衡状态极有可能是不存在的。因此，热寂说面临着新的更为严峻的考验。

对于热力学第二定律的界定，中国学者张邦固在《恒星起源动力学》一书中给予了科学的论证。结论是：由忽略了引力的宏观系统总结出来的熵增原理不适用于以引力为主的宇观系统。因为，在宏观系统，引力被忽略，无规热运动占主导地位，体现系统无规运动的熵就倾向增加；在宇观系统，引力占主导，粒子的运动趋向有序，熵就要减少。

至此，热寂说的基石已经去除，没有必要再杞人忧天。

概括论述一下：宇观世界的熵与宏观世界的熵不同。

（1）宏观系统中，引力被完全忽略了。例如，在热机中起重要作用的气体系统中，气体分子之间的作用基本上被忽略了。在比较细致的研究中，也只是考虑分子之间的范德瓦耳斯力。大家知道，这也是一种电磁相互作用。

（2）在宇观系统中，基本上只有引力起作用。

（3）孤立宏观气体系统倾向于膨胀。一个气体系统，只要打开它的壁，它就会膨胀。而孤立宇观气体系统会在自身引力作用下收缩，早期恒星就是这样形成的。对此有兴趣的读者可以看看《恒星起源运动学》（科学出版社，1994）。

可以看到，正是两种系统的区别导致了两种系统的熵发生截然相反的变化。

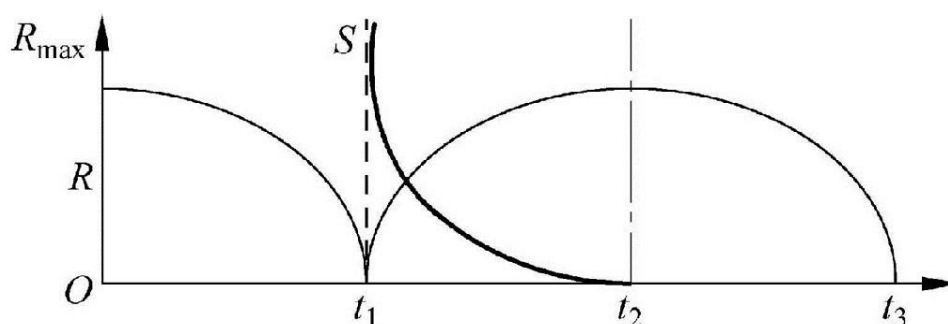
（1）宏观气体系统倾向于膨胀。膨胀之后，空间大了，对应的微观状态就会增多，熵就会变大。而宇观孤立气体系统会收缩，相应地它的熵会减少。

（2）宏观气体系统中引力被忽略了，起主要作用的是热作用。这是使运动更加混乱的作用，所以熵会增加。宇观孤立气体系统中引力起主要作用，引力是使粒子运动更加有序的作用，系统的熵会减少。

实际的计算表明，在早期恒星形成的过程中，它的熵确实是在不断减少。

我们宇宙的熵在不断减少：

我们宇宙完全“浸泡”在背景辐射之中，是一个热平衡态。它的熵是容易计算的。在下图中：大爆炸发生时（ t_1 ），我们宇宙的半径（细线）最小，熵（粗线）最大。之后，随着宇宙的膨胀，熵不断减少，直到宇宙膨胀结束（ t_2 ），背景辐射消失，系统不再是热力学状态，运动不再混乱，系统的熵等于零。



我们宇宙的熵 S （粗线）

对于宇宙膨胀的结果及未来的趋势，必须强调宇宙微波背景辐射的光子平衡态性质，这是客观的，最根本的。因此它定性地决定了，我们宇宙是引力束缚态，不会永远膨胀；根据天文观测的背景辐射温度、宇宙密度和宇宙半径数据，计算我们宇宙的总能量，可以发现宇宙总质量小于宇宙总静止质量，这就定量地说明，我们宇宙是引力束缚态。

由背景辐射温度和观测到的宇宙密度可以得到，光子数是粒子数的100亿倍。进而推断，宇宙的整体运动由背景辐射决定。

由观测到的众星系退行表明，背景辐射已经膨胀了100多亿年。可以得出，光子平衡态必然膨胀。这就是大爆炸的原因。

下面分析宇宙从目前的膨胀到将来收缩的转折过程。希望能找到发

生这种转折的原因。

和多数讨论宇宙整体运动的文章一样，我们也采用抹平了的宇宙模型。具体地说，就是不考虑形成恒星、星系等局部的运动。在光子平衡态的背景下，氢原子、中子等粒子均匀地分布在宇宙中。由于粒子的相对数极少，不妨形象地叫它做“光子清汤”。

随着宇宙继续膨胀，背景辐射的温度会进一步降低。当它由目前的2.735开降到0开时，宇宙便会转入收缩。理由如下：

1) 光子平衡态必然膨胀。前面已经介绍过，天文观测表明我们的宇宙在膨胀着。也就是说，这是一个事实。并且，这是不可能偶然发生的。只能把它看成是系统本身固有的性质。

(1) 重子等静止质量不为零的粒子系统可能收缩，例如，恒星就是一团气体收缩而成的。

(2) 宇宙整体的热力学性质由光子平衡态决定。

①在过去100多亿年以上的时间里，光子平衡态一直在膨胀着。这不仅是大爆炸理论的内容，也是我们观测到的客观事实，我们观测到的遥远的星系退行正是表现了几亿、几十亿……年前宇宙光子平衡态膨胀的信息。

②这种膨胀是光子平衡态自身的特性。它不同于弹片飞向四面八方的“膨胀”。弹片的“膨胀”源于起初炸药爆炸，它们飞向四面八方，均远离它们的质心。宇宙光子平衡态中任何一块空间，必然包含着奔向各方的光子，既有背离宇宙质心的，也有朝着质心的。它也不同于超新星类型的爆炸。超新星爆炸后形成的星云是一层气体壳。这层气体壳与残留

的核之间有着广袤的密度相对小的空间。光子平衡态中物质是三维均匀的。

③由于系统中光子数占绝对多数，与光子之间的相互作用比较，粒子所参与的相互作用对系统性质的影响可以忽略。

④与其他任何粒子不同，光子永远以光速运动，无论相对于任何参考系。因此，光子的平衡态也必然有与众不同的性质。只有到了热力学温度零度（0开），光子平衡态消失了，宇宙膨胀才会停止。所以说，必然膨胀是光子平衡态的固有性质。

2) 从能量转换的角度来看。宇宙膨胀过程是宇宙物质的动能逐渐转化为引力势能的过程。宇宙膨胀到最大时，应该是其动能最小（零）之时。光子平衡态正是宇宙动能形态的主要体现。

3) 从运动形态上看。大爆炸后，没有与重子脱耦之前，光子数与重子数大致相同。单个粒子与单个光子的平均动能大致相等。这种热平衡态可以形象地叫做“光子浓汤”。宇宙不断膨胀，“光子浓汤”变成了“光子清汤”。宇宙一直处于热平衡态之中。所有粒子和光子均在热运动，系统的熵不等于零。相反，在收缩过程中，所有粒子均同步地奔向质心，是完全有序的运动。系统的熵恒为零。

当宇宙膨胀到 $R=R_{\max}$ ，温度降到了热力学温度零度（0开），宇宙微波背景辐射彻底消失了。各星系不再退行，处于相对静止状态。这样，宇宙将不会进一步膨胀。在重子之间引力的作用下，宇宙将进入收缩阶段。

通过具体的计算，有 $R_{\max}=2.13\times 10^{10}$ 光年。

我们看到，此时宇宙已经处于膨胀的尾声，继续膨胀约6%以后，它将不再膨胀。在引力作用下开始收缩。在均匀分布条件下得到的收缩是同步的。

当宇宙中众粒子同步到达宇宙质量中心附近一个相对小（大约 $R_{\min}=4.4377\times 10^{13}$ 米。为了对这个宇宙的最小半径有一个感性的认识，我们把地球到太阳的平均距离列在下面：1天文单位 $=1.496\times 10^{11}$ 米。也就是说，它比日地距离还大了近300倍！这样的宇宙球怎么会是一个点？）的区域时，会发生剧烈碰撞，从而激发下一个光子平衡态。

感兴趣的读者可见《宇宙中航行》（张邦固著，知识产权出版社，2014年2月）。

5.2.5 黑洞理论

最早涉及黑洞这一类天体的科学家，既不是爱因斯坦，也不是霍金，而是一位18世纪的剑桥大学学监。这位学监的名叫约翰·米歇尔，他在《伦敦皇家学会哲学学报》上发文指出，对于一定质量的星体而言，其周长越小，表面的引力越强，粒子逃离星体的引力作用就越强。根据这个前提，他提出，星体存在一个临界周长，对于粒子而言其逃逸速度为光速。那些周长小于临界周长的星体，光粒子无法从表面逃逸，星体便成了黑暗星。这个黑暗星，也便是我们现在称之为黑洞的天体。

1915年爱因斯坦相对论发表后，相对论方程式成为众多科学家的研究对象。1916年，德国物理学家卡尔·史瓦西（1873—1916）找到了广义相对论球对称引力场的严格解，被称为“史瓦西度规”。在解释这一概念时，他提出，在距离极高密度天体或大质量天体的中心的某一距离

上，逃逸速度相当于光速。简言之，就是在此距离内的任何物质或者辐射，都无法逃逸这类星体。这便是史瓦西半径，史瓦西半径上的点形成的球面被称为“视界”。

1928年，一位来自印度的18岁天才钱德拉塞卡（公元1910—1995年）在前往英国的途中计算出了一个极限值：恒星在耗尽自己所有的燃料之后，多大的质量使它能够对抗自己的引力而不坍缩。其意思是，恒星在生命的最后阶段，引力使其缩小，粒子变得十分致密，粒子间相互排斥并使得恒星存在膨胀倾向。这与粒子间的引力作用相反，两种相反的力便决定了恒星的最终归宿，而其关键作用的因素，便是恒星的质量。

钱德拉塞卡计算出，质量超过太阳1.44倍的恒星，在最后阶段将无法维持自身大小以对抗引力。这便是后来著名的钱德拉塞卡极限。他在1935年宣读自己的论文时，提出了这个理论。不过，结果却是爱丁顿将其讲稿撕成两半。爱因斯坦也发文反对这个理论。因为按照这个理论，质量超大的恒星甚至可能坍缩为一个点。差不多30年后，一些天文学家才接受了这个理论。

1939年，奥本海默（1904—1967）与他的研究生斯奈德在《物理学评论》上发表文章，他们指出，当恒星收缩到某一临界半径时，其表面引力大到使光线严重偏折，直至无法逃离。当时，奥本海默的这一理论并未得到重视，他本人随后也投入到了原子弹的制造中。直到20世纪60年代，现代天文学技术（尤其是射电望远镜的应用）的发展使得研究宇宙大尺度问题成为新的热点，奥本海默的工作才被天文学家重新发现并推广。

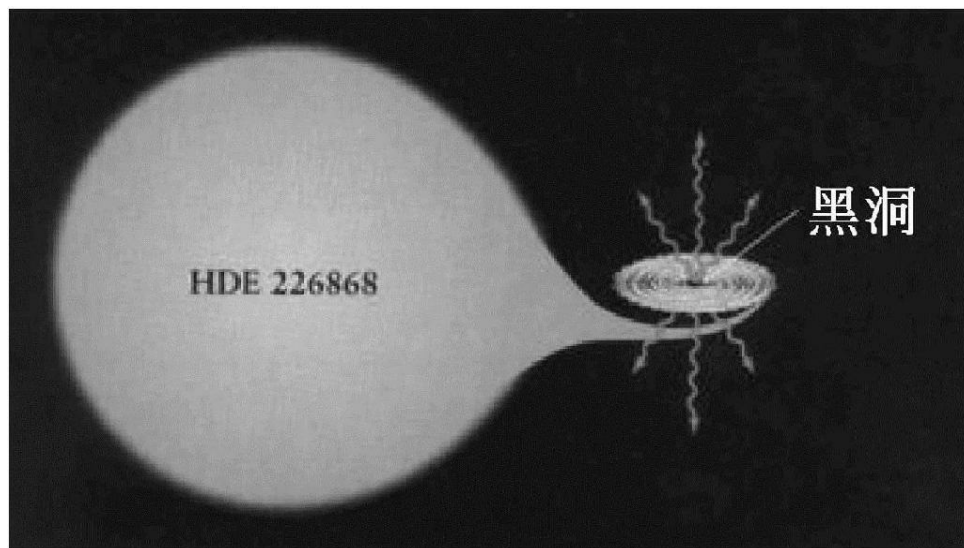
1964年，彭罗斯证明了所有内爆的恒星（也就是暗星）一旦形成事

件视界（视界内的物质，将无法对视界外的事物产生任何影响，视界外的物质也无法观测到视界内的事物，实际就是黑洞的不可逃逸范围。）就必然会成为奇点。在此之后，黑洞理论有了长足的发展。1967年，美国物理学家约翰·阿奇博尔德·惠勒（1911—2008）正式提出了“黑洞”这一名词。1969年，彭罗斯再次证明可以提出黑洞的能量，并提出物理定律不允许裸奇点生成的宇宙监督假设。裸奇点是理论中没有被视界包围住的引力奇点，光和粒子都能够从中逃逸，使得我们能够观测奇点附近的时空弯曲。假如时空出现了裸奇点，物理学定律将失去描述未来事件的力量，这便是宇宙监督假说。1971年，霍金提出了原始黑洞诞生于宇宙早期的理论。1974年，他又提出黑洞具有热辐射的理论，“这个重要的发现告诉我们黑洞没有原来认为的那么黑”。

黑洞附近由于巨大的引力作用，存在强烈的时空弯曲，能够使光线产生偏折现象。在这种情况下，即便是有一部分光进入事件视界后无法逃逸，但仍有部分光线绕过黑洞而到达地球。因此，在我们的直接观测中，黑洞似乎并不存在，这便是黑洞迄今为止无法通过直接观测确认的原因。

但这并不意味着黑洞无法检测。“正如约翰·米歇尔在他1783年的先驱性论文中指出，黑洞仍然将它的引力作用到周围的物体上。”天文学家观测到许多双星系统，其中有一些系统，一颗可见的恒星环绕着另一颗不可见的恒星运动，最著名的是天鹅座X-1双星系统。不过仅仅依靠这一点，并不能确定黑洞的存在，因为不可见的恒星有可能只是由于星等较低观测不到。这时就需要借助万有引力与钱德拉塞卡极限的理论了。从不可见星对可见星的摄动情况，我们可以分析出不可见星的引力与大小；根据可见星的运行轨道，利用双星运行模式的万有引力方程，我们可以得出不可见星的最小可能质量。而计算表明，天鹅座X-1双星

系统中的不可见星的最小质量约为太阳的6倍，且这个伴星的引力奇大，体积却非常小。根据钱德拉赛卡极限，超过太阳质量1.44倍的恒星，将有可能坍缩为中子星。而质量达到太阳6倍的恒星，只可能是黑洞。



天鹅座X-1双星系统是人类通过此种方式发现的第一个可能的黑洞，发现时间为1971年。自此之后，便不断有黑洞被发现，黑洞的研究也更进一步。到2011年11月，借助哈勃空间望远镜，科学家首次拍摄到围绕遥远黑洞周围的盘状构造。



依照目前的探测手段，黑洞能够探测到的物理性质只有三个：质量、电荷、角动量。这三个量是无法变为电磁辐射的守恒量，确定这三个量后，便确定了黑洞的大小、形状与其他一切信息。这个理论被称为“无毛定理”，也可以用一句格言总结：“黑洞没有毛。”“毛”代指黑洞除了三个可探测属性外的大量信息，“无毛”表明我们无法探测这些信息。“所以，无毛定理还意味着，有关这个天体的非常大量的信息，在黑洞形成时消失了。”

对于黑洞结构的研究，目前以霍金的理论最为学界接受。在《时间简史》中，霍金花了整整两章的篇幅，详细描述了目前黑洞研究的最新理论成果。

之后的三年，霍金试图将量子力学引入黑洞理论中，论证了“霍金辐射”。然而霍金认为，即便黑洞将自身能量以黑洞辐射的形式辐射出去，自身渐渐蒸发，失去质量，这些辐射却无法反映之前的所有信息，

所有的信息将随着黑洞的蒸发而消散。而根据量子理论，这样的信息永远不可能消失。在这里，霍金的理论与量子力学发生了冲突，这便是著名的“黑洞悖论”。实际上霍金并未将“黑洞悖论”著入《时间简史》中，或许是因为在那是他自己也不甚确定。直到2004年，霍金在一次国际会议中表示，自己所坚持的“黑洞悖论”是错误的。

在《时间简史》中，霍金还提到一类奇特的黑洞——太初黑洞。太初黑洞也叫“原初黑洞”，霍金于1971年就预言了太初黑洞的存在。这种黑洞并非由恒星坍缩而成，而是在宇宙大爆炸的初期，某些区域密度非常大，以至于宇宙膨胀后这些区域的密度仍然大到可以形成黑洞。这些黑洞的质量并不大，只有10亿吨左右（一座大山的质量）；体积很小，甚至小于原子核大小。而到目前为止发现的最大的黑洞，质量大概是太阳的100亿倍。银河系中心的黑洞，也有太阳的400万倍。但大多数黑洞的质量为太阳的数倍。大质量黑洞通常位于星系的中心，维持着整个星系的运转。

霍金在提出“霍金辐射”理论时曾提出，质量大的黑洞，温度越接近热力学温度零度，蒸发越慢；质量小的黑洞，温度稍高，蒸发速度越快。质量很小的太初黑洞可能已经蒸发或者即将蒸发，天文学家们试图侦测太初黑洞最后蒸发时发射的伽马射线。

霍金对黑洞理论的贡献，不仅仅在于研究成果的不断深入，更在于研究方法与思维方式的更新。他身残志坚的奋斗精神令人赞叹。30年前，黑洞对大多数科学家来说是一个陌生的名词。而现在，稍有文化的人都对黑洞有些许的了解。但是，他提倡奇点说，在奇点时一切已知规律都不遵守；时间、空间都以此为起点；以及黑洞悖论等都是错误的，我们应当分清。

史蒂芬·威廉·霍金，1942年1月8日出生于英国牛津。毕业于牛津大学和剑桥大学，并获剑桥大学博士学位。英国剑桥大学应用数学与理论物理学系物理学家，1979年至2009年任卢卡斯数学教授。牛顿也曾任此职，是人类历史上最伟大的教授职位。霍金在21岁时不幸患上了会使肌肉萎缩的卢伽雷症，所以被禁锢在轮椅上，



霍金

现在只有三根手指可以活动。疾病已经使他的身体严重变形，后来他因患肺炎做了穿气管手术，被彻底剥夺了说话的能力，演讲和问答只能通过语音合成器来完成。当时医生预测他最多活2年，但至今他依然活跃在科学界。1973年，他考察了黑洞附近的量子效应，发现黑洞会像天体一样发出辐射，其辐射的温度和黑洞质量成反比，这样黑洞就会因为辐射而慢慢变小。而其温度却越变越高，最后以爆炸而告终。黑洞辐射或霍金辐射的发现具有极其重要的意义，它将引力、量子力学、统计力学统一在了一起。

1974年以后，他的研究转向了量子引力论。虽然人们还没有得到一个成功的理论，但它的一些特征已被发现。例如，空间—时间在普朗克尺度（ 10^{-33} 厘米）下不是平坦的，而是处于一种粉末的状态。在量子引力中不存在纯态，因果性受到破坏，因此使不可知性从经典统计物理学、量子统计物理提高到了量子引力的第三个层次。

1980年以后，霍金的兴趣转向了量子宇宙论，提出了能解决宇宙第一推动问题的无边界条件。2004年7月，他承认了自己原来的“黑洞悖论”观点是错误的。霍金认为他一生的贡献是在经典物理的框架里，证

明了黑洞和大爆炸奇点的不可避免性，黑洞越变越大，但在量子物理的框架里，他指出，黑洞因辐射而越变越小，大爆炸的奇点不断被量子效应所抹平，而且整个宇宙正是起始于此。但是，奇点说是根本错误的。

20世纪70年代时，霍金将量子力学应用于解释黑洞现象。但是在之后的30年中，用量子力学解释整个宇宙已经变得更加困难了。霍金想找到一套可以完美解释整个宇宙现象的理论，来说明137亿年诞生后直到现在的宇宙，但是多年过去了，他仍然没有得出结论。按照他的量子力学理论，宇宙诞生是大爆炸产生的，这是一个被压缩的无限小却具有超大重力的物质（也可以理解成密度无限大）爆炸的产物。量子力学的理论范畴不能够解释这一个过程是如何进行的，为什么会这样。霍金说“那必须有一套可以描述小规模重力的理论”。最新的科学突破是霍金的同事，伦敦玛丽皇后学院的麦克·格林参与建构的超弦理论，简称为“弦论”，这理论指出所有粒子和自然力量，其实都是在震荡中的像弦一样的微小物体，解决了霍金一直想努力解答的重力问题，不过这个理论建立在宇宙必须有9、10个甚至是大于11个的维度中，而人类身处的三维世界可能仅仅是真正宇宙的其中一个膜.....但是，弦论至今没有得到证实。

2014年1月24日，霍金教授再次以其与黑洞有关的理论震惊物理学界。他在日前发表的一篇论文中承认，黑洞其实是不存在的，不过“灰洞”的确存在。在这篇名为《黑洞的信息保存与气象预报》的论文中，霍金指出，由于找不到黑洞的边界，因此黑洞是不存在的。经典黑洞理论认为，黑洞外的物质和辐射可以通过视界进入黑洞内部，而黑洞内的任何物质和辐射均不能穿出视界。但量子力学理论表明，能量和信息是可以从黑洞中出来的。但霍金的最新“灰洞”理论认为，物质和能量在被黑洞困住一段时间以后，又会被重新释放到宇宙中。他在论文中承认，

自己最初有关视界的认识是有缺陷的，光线其实是可以穿越视界的。当光线逃离黑洞核心时，它的运动就像人在跑步机上奔跑一样，慢慢地通过向外辐射而收缩。

霍金同时指出，对于这种逃离过程的解释需要一个能够将重力和其他基本力成功融合的理论。在过去近100年间，物理学界没有人曾试图解释这一过程。

对于霍金的“灰洞”理论，一些科学家表示认可，但也有人持怀疑态度。美国卡夫立理论物理研究所的理论物理学家约瑟夫·波尔钦斯基指出，根据爱因斯坦的重力理论，黑洞的边界是存在的，只是它与宇宙其他部分的区别并不明显。霍金的理论缺乏观测事实的支持和验证，这是其致命的弱点。

5.3 天体观测的进展

5.3.1 射电天文学的诞生

在第4章中已经对射电天文望远镜作了较详尽的讲述，下面从射电天文学的角度，作一重点介绍。

1. 央斯基的发现

1931年至1932年期间，捷裔美籍无线电工程师央斯基在贝尔电话实验室工作。他用长30.5米、高3.66米的旋转天线阵检测无线电通信的各种干扰因素时，探测到一种来源不明的稳定的天电噪声。1935年，人们确认这是来自银河系中心的射电辐射。

2. 雷伯的经典式射电望远镜

1937年，美国无线电工程师雷伯建造了9.45米口径的抛物面天线，这是第一台经典式射电望远镜。1940年，他确认了央斯基的发现，并绘出了对应于银心的射电源的等强度线。论文发表于《天体物理学杂志》。

由于央斯基和雷伯的开创性工作，射电天文学得以诞生，这是天文学发展史上又一次新的飞跃，使得人类能够接收来自宇宙的无线电波段的信号，极大地扩展和深化了人类认识宇宙的视野。从此，许多前所未有的发现接踵而来。

5.3.2 银河系结构的射电探测

1. 21厘米微波辐射的理论预言

1944年，荷兰天文学家范得胡斯特（1918—）从理论上预言太空中的中性氢区发射波长为21.2厘米的微波辐射。

2. 21厘米微波辐射的探测

1951年，美国的尤恩和珀塞尔首次观测到来自银河系的21厘米谱线信号。接着荷兰的穆勒和奥尔特，以及澳大利亚的克里斯琴森等也观测到了。

1958年，荷兰和澳大利亚的天文学家联合探测，绘制了银河系的中心氢分布图。图上清楚地显示了银河系具有旋涡结构，并发现了几条旋臂。



5.3.3 20世纪60年代的四大天文发现

1. 类星体的发现

1960年，美国天文学家马修斯和桑德奇等发现射电源3C 48对应的

光学像类似于一颗恒星，其光谱有一些奇怪的发射线。后来发现它的主要发射线实际上是红移量达0.367的氢线。

1963年，美国天文学家M.施米特拍摄了射电源3C 273中恒星状天体的光谱，主要发射线是红移达0.158的氢线。

这类天体称类星射电源，通称“类星体”。

2. 微波背景辐射的发现

1964年7月至1965年4月，美国科学家在波长7.35厘米的微波段进行探测，发现有相当于 3.5 ± 1 开的剩余噪声。噪声是各向同性的，而且没有季节性变化。

1965年下半年，普林斯大学物理教授迪克领导的小组在3.2厘米波长处也观测到了3K微波背景辐射。

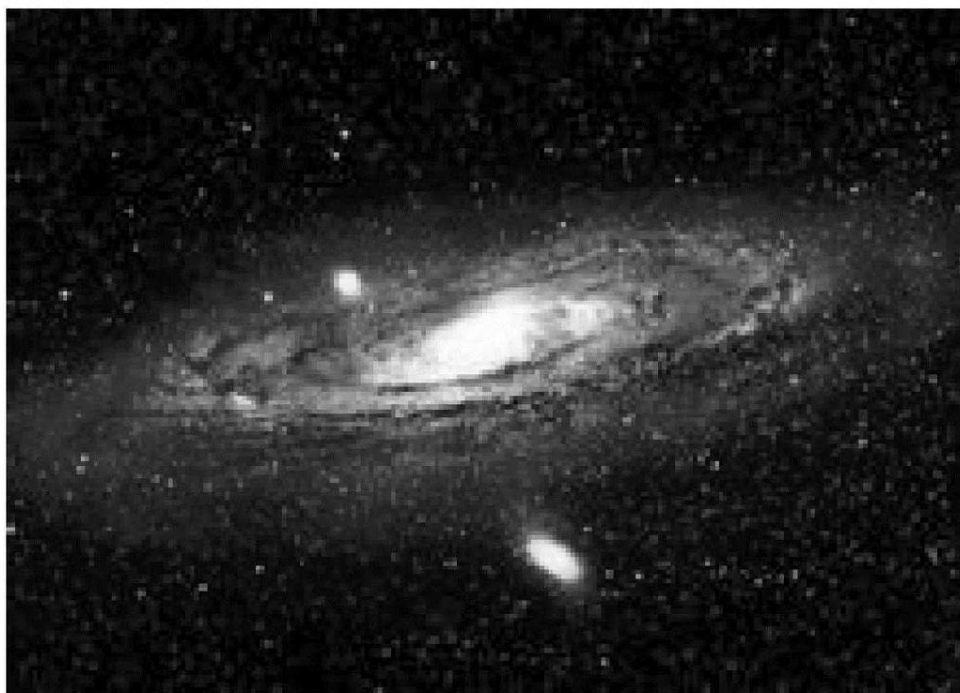
3. 射电脉冲星的发现

1967年7月，在英国天文学家休伊什的主持下，剑桥大学卡文迪许实验室建成了一架时间分辨率很高的射电望远镜。望远镜工作波长为3.7米，天线是具有2048个的偶极子的天线阵，占地面积18 212平方米。1967年10月，贝尔发现记录纸带上有奇怪的脉冲信号。

1967年11月28日，记录表明这个位于狐狸座的射电源以1.337秒、极精确周期的脉冲形式辐射出射电波。这类发出射电脉冲的射电源称为射电脉冲星。

1974年，休伊什获诺贝尔物理奖。

1939年，美国原子物理学家奥本海姆预言并建立了第一个中子星模型。1968年，美国天文学家戈尔德指出射电脉冲星的本质就是高速自转的中子星。



4. 星际有机分子的发现

1963年和1968年12月，美国先后发现星际羟基分子（OH）和星际氨分子（NH₃），不久又发现了星际水分子（H₂O）。

1969年3月，美国一个小组用直径43米的射电望远镜在射电源人马座A和人马座B2背景上发现星际甲醛分子（H₂CO）。这是人类发现的第一个星际有机分子。之后，又发现了许多星际有机分子。星际有机分子的发现为宇宙化学和生命起源的研究揭开新的一页。

20世纪60年代的四大天文发现标志射电天文学已经成熟，也表明射电天文学为宇宙、天文学开辟了全新的研究领域。

5.3.4 射电天文学的新进展

1. 甚长基线射电干涉测量

1965年，美国佛罗里达大学的一个研究小组首次使用了各自独立的磁带记录干涉仪探测木星的射电爆发区域，获得0.1"的分辨率。

当前甚长基线射电干涉测量（VLBI）的分辨率已达0.000 2"。

2. 综合孔径射电望远镜

20世纪50年代，英国天文学家赖尔首先提出制造综合孔径射电望远镜的设想。1963年，他完成了两个天线最大间距为1.6千米的综合孔径射电望远镜。它的研制成功使射电源的成像成为可能。这是射电天文技术的一项重大突破。1971年，赖尔主持制造英国剑桥大学的“五千米阵”。1974年，赖尔荣获诺贝尔物理学奖。

当前最大的综合孔径射电望远镜是美国的甚大阵（VLA）。

3. 毫米波和亚毫米波天文学

星际分子的谱线波长多在1~10毫米的波段和0.35~1毫米的亚毫米波段。随着空间探测能力的提升，形成射电天文学中新分支——毫米波和亚毫米波天文学。

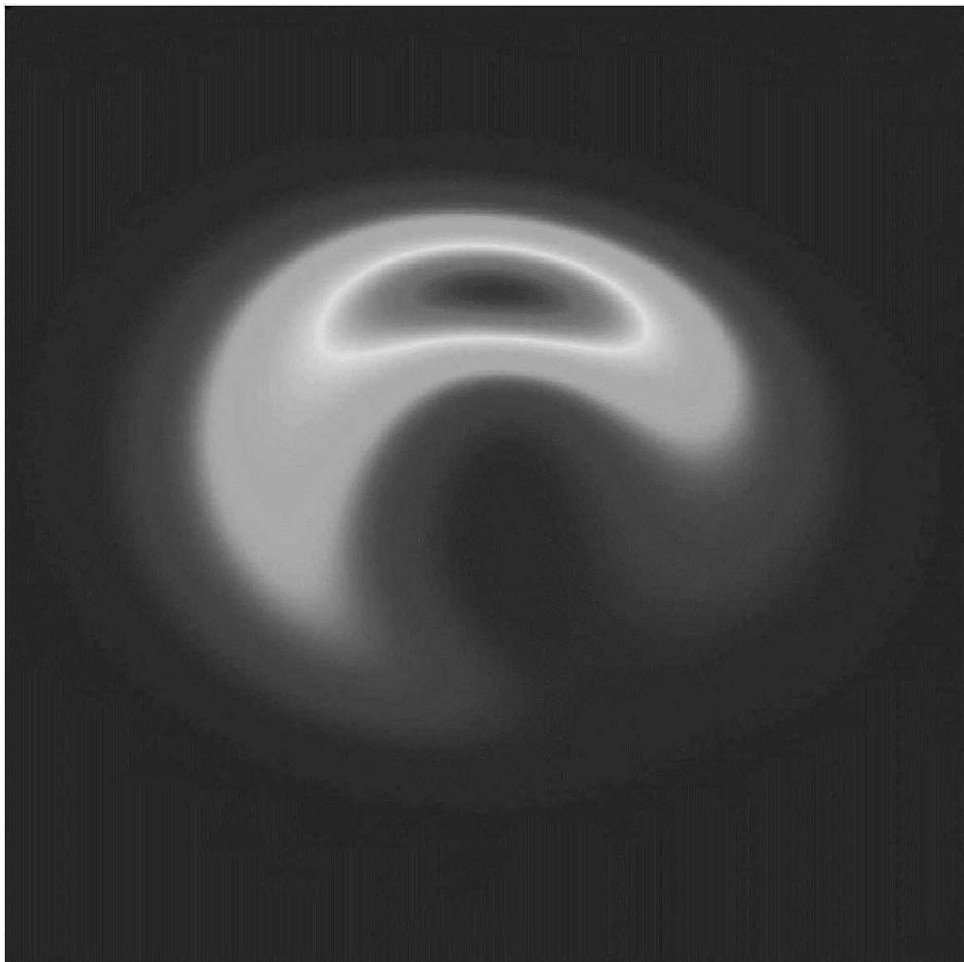
4. 全波段天文学

天体发射着 10^{-12} ~ 10^8 厘米范围的电磁波，但只有可见光和两侧的近紫外、近红外、约1毫米至30米的射电波，以及红外波段的几小段波长区间内的辐射才能到达地面。

20世纪40年代以来，首先射电天文学问世，随之红外天文学、紫外天文学、X射线天文学、 γ 射线天文学等一系列新的天文学科也产生了，这些新学科大部分与航天科技及空间探测的进展紧密相关。由此，宇宙天文学从光学天文学进入全波段天文学。

5. 发现更多、更大的黑洞

2011年初，科学家宣称，在我们所在的银河系中央存在一个黑洞。其质量比太阳的大400万~500万倍。



2015年初，中国北京大学发布：以中国天文学家为主的科研团队发现了一颗距离地球128亿光年的、430万亿倍太阳光度的超亮类星体，其

中心黑洞质量约为120亿倍太阳质量，比先前发现的同时期黑洞质量的总和还要大1倍。德国科学家也宣布了同样的发现。

5.4 宇航时代到来

5.4.1 空间探测手段的发展

(1) 气球。1783年，法国蒙哥尔费兄弟研制成的热空气气球首次升空，后来又用氢和氦代替热空气。

现代的气球最高可升到52千米高，载重5吨，留空时间10年。可以把99.9%的大气留在气球之下。这一阶段的天空中几乎没有水汽和二氧化碳，对红外观测有利。

(2) 火箭。由中国人最早发明。北宋开宝三年（970年）冯继升和岳义方两人作了首次成功试验。

(3) 现代宇宙航行的奠基者

俄国的齐奥尔科夫斯基（1857—1935）为利用火箭作宇宙飞行奠定了理论基础。

1926年，美国物理学家戈达德在马萨诸塞州发射了第一枚现代意义上的火箭。

1923年，德裔科学家奥伯特出版《飞往星际空间的火箭》一书，论述了火箭飞行理论和火箭结构，后来又参加了V—2火箭的研制。

齐奥尔科夫斯基、戈达德和奥伯特被公认为现代宇宙航行的奠基者。

20世纪40—50年代用高空火箭飞出大气外作空间控制探测。

（4）人造地球卫星和宇宙飞船

1957年10月4日，苏联发射了第一颗人造地球卫星“卫星1号”。11月3日又发射了第二颗——“卫星2号”。

1958年1月31日，美国发射了人造地球卫星——“探险者1号”。

人造地球卫星的上天标志着人类进入了宇航时代，也为宇航空间探测开辟了广阔的道路。地球是人类的摇篮、宝贵的家园。但人类已长大，必然会飞往太空，去探寻无限神秘的未知世界，去破解萦绕心中的各种谜团。尽管征途上充满艰险，会有伤痛，仍阻挡不住人类前进的步伐。



5.4.2 开展空间探测

1. 地球辐射带的发现

人造地球卫星发射后的第一项天文学成就是通过对地球周围空间的探测发现了地球上空的两个带电粒子区——辐射带。

1959年，美国天文学家范艾伦根据“探险者1号”、“探险者4号”和“先驱者3号”等探测器的探测结果发现了地球上空两个辐射带：内、外范艾伦带。

2. 对月球的探测

（1）苏联对月球的探测

由“月球”号飞船系列实现。

①1959年10月4日，“月球”3号首次拍摄了月球背面的图像。

②1966年2月3日，“月球”9号首次在月球上软着陆。

③1970年9月，“月球”16号首次挖取月球样品，并返回地球。

④1970年11月，“月球”17号；1973年1月“月球”21号，各携带一辆月行车，在月面巡视考察。

（2）美国对月球的探测

早期分别由“徘徊者”、“勘测者”、“月球轨道环行器”和“阿波罗”4个飞船系列实现各阶段任务。

①1961年8月至1965年3月，发射了“徘徊者”系列9个飞船，仅最后3个实现硬着陆。

②1966年5月至1968年1月，“勘测者”系列中5个成功实现软着陆。

③1966年5月至1967年8月，“月球轨道环行器”和5艘绕月飞船，为“阿波罗”飞船选择着陆点。

④1965年2月至1972年12月，“阿波罗”系列的共17艘飞船。前6艘无人，7~10号为载人试验。

1969年7月16日，“阿波罗”11号发射，7月20日两名字航员登陆于月面静海。



之后，“阿波罗”12、14、15、16、17均载人登月成功，在月面进行多项科学活动，采回月岩、月壤样品400多千克。

（3）中国对月球的探测

2007年11月26日，中国正式公布“嫦娥一号”卫星传回的第一幅月面

图像。2009年3月1日，“嫦娥一号”卫星在控制下成功撞击月球。为我国月球探测的一期工程，画上了圆满句号。



“嫦娥二号”于2010年10月1日在西昌卫星发射中心发射升空，并获得了圆满成功。

2020年前，我国的月球探测工程以无人探测为主，分三个实施阶段。

“绕”：2004年至2007年（一期），研制和发射我国首颗月球探测卫星，实施绕月探测。这一阶段主要任务是研制和发射月球探测卫星，突破绕月探测关键技术，对月球地形地貌、部分元素及物质成分、月壤特性、地月空间环境等进行全球性、整体性与综合性的探测，并初步建立我国月球探测航天工程系统。

“落”：2013年前后（二期），进行首次月球软着陆和自动巡视勘

测。主要任务是突破月球软着陆、月面巡视勘察、深空测控通信与遥控操作、深空探测运载火箭发射等关键技术，研制和发射月球软着陆探测器和巡视探测器，实现月球软着陆和巡视探测，对着陆区地形地貌、地质构造和物质成分等进行探测，并开展月基天文观测。

“回”：2020年前（三期），进行首次月球样品自动取样返回探测。主要任务是突破采样返回探测器小型采样返回舱、月表钻岩机、月表采样器、机器人操作臂等技术；在现场分析取样的基础上，采集关键性样品返回地球，进行实验室分析研究；深化对地月系统的起源与演化的认识。

在“绕”、“落”、“回”均成功实现以后，我们才能进行下一步的人登上月球的计划。

对月探测使宇宙、天文学的研究方法产生了历史性变化，由局限于观测方法，发展为了可以同时采用实验方法。

3. 对水星和金星的探测

对水星探测：

1973年11月3日，美国发射“水手”10号。1974年2月5日，探测金星（距离5800km）后，受金星引力作用后加速，进入与水星轨道在远日点相切的绕日运动轨道。3月29日近距离探测水星（距离703km）。9月21日再次近距离探测水星。次年3月16日最后一次探测水星（距离327km）。

2004年，美国发射“信使号”火星探测器，并于2011年进入火星环绕轨道。

对金星探测：

方式：逼近飞行、硬着陆、绕转飞行。

苏联：“金星”号飞船系列，1961年至1981共14艘。

美国：

第一艘1961年5月20日，“金星”1号飞越金星。

首次硬着陆，1967年10月18日，“金星”4号。

首次软着陆，1970年12月15日，“金星”7号。

首批金星表面照片，1975年，“金星”9号、10号。

首次绕金星飞行，1978年12月，“先驱者—金星”1号。

4. 对火星的探测

苏联：“火星”号飞船系列，1962年11月至1973年8月，共7艘。

美国：“水手”4、6、7、9号飞船；“海盗”1、2号飞船。

首次软着陆，1971年12月，“火星”3号。

1976年7月20日“海盗”1号，1976年9月3日“海盗”2号，软着陆分析土壤样品，未发现生命迹象。

5. 对类木行星的探测

美国两个系列的飞船：“先驱者”号和“旅行者”号。

1972年3月3日发射“先驱者”10号，1973年12月4日飞越木星。

1973年4月5日发射“先驱者”11号，1974年12月3日飞越木星。1979年9月飞越土星。

1977年8月20日发射“旅行者”1号，1979年3月5日飞越木星，1980年8月飞越土星。

1977年9月5日发射“旅行者”2号，1979年7月飞越木星，1981年8月飞越土星，1986年1月飞越天王星，1991年飞越海王星。

航天科技及太空探测对宇宙科学发挥着越来越重要的作用。

5.4.3 载人航天

1961年4月12日，苏联宇航员加加林（1934—1968）成为世界飞入太空的第一人。

1965年3月18日，苏联宇航员列昂诺夫走出“上升”2号飞船，离船5米，停留12分钟，首次实现人类航天史上的太空行走。

苏联宇航员科马洛夫，1967年4月24日乘“联盟”1号飞船返回地面时，因降落伞未打开，成为第一位为航天殉难的宇航员。

1969年1月14至17日，苏联的“联盟”4号和5号飞船在太空首次实现交会对接，并交换了宇航员。

1969年7月20日，美国宇航员阿姆斯特朗乘坐“阿波罗”11号飞船，成为人类踏上月球的第一人。

1971年4月9日，苏联“礼炮”1号空间站成为人类进入太空的第一个空间站。两年后，美国将“天空实验室”空间站送入太空。



1981年4月12日，世界第一架航天飞机——美国的“哥伦比亚”号航天飞机发射成功。

1981年4月21日，美国成功发射并返回世界上首架航天飞机“哥伦比亚”号，使可重复使用的天地往返系统梦想成真。

1984年2月7日，美国宇航员麦坎德列斯和斯图尔特不拴系绳离开“挑战者”号航天飞机，成为第一批“人体地球卫星”。

1985年7月25日，王赣骏乘“挑战者”号航天飞机进入太空，成为第一位华裔宇航员。

1986年1月28日，“挑战者”号航天飞机起飞时发生爆炸，7位宇航员全部遇难，成为迄今最大的一次航天灾难。

1986年2月20日，苏联发射“和平”号空间站，服役已经超期8年，是目前最成功的人类空间站。

1993年11月1日，美、俄签署协议，决定在“和平”号空间站的基础上，建造一座国际空间站，命名为“阿尔法国际空间站”。

俄罗斯的波利亚科夫于1994年至1995年间在“和平”号空间站上连续停留438天，成为在太空停留时间最长的男宇航员；而美国的露西德于1996年在“和平”号上停留了188天，成为在太空停留时间最长的女宇航员。

1995年6月29日，美国“亚特兰蒂斯”号航天飞机与俄罗斯“和平”号空间站第一次对接，开始了总计9次的航天飞机与空间站的对接，为建造国际空间站拉开了序幕。

2003年2月1日，“哥伦比亚”号航天飞机在返回途中解体，机上7人全部丧生，迫使美国停止航空飞机发展计划。

2003年10月15日，中国首位宇航员杨利伟乘“神舟五号”进入太空。



中国航天史是从1956年2月开始的，当时著名科学家钱学森向中央提出《建立中国国防航空工业的意见》。

1960年2月19日，中国自行设计制造的试验型液体燃料探空火箭首次发射成功。



1970年4月24日21时31分，中国“东方红”一号飞向太空。这是中国发射的第一颗人造地球卫星。

1971年3月3日，中国发射了科学实验卫星“实践一号”。卫星在预定轨道上工作了8年。

1971年9月10日，我国的洲际火箭首次飞行试验基本成功。

1975年11月26日，中国发射了一颗返回式人造卫星。卫星按预定计划于29日返回地面。

1984年4月8日，中国第一颗地球静止轨道试验通信卫星发射成功。16日，卫星成功地定点于东经125°赤道上空。

1986年2月1日，中国发射一颗实用通信广播卫星。20日，卫星定点成功。这标志着中国已全面掌握运载火箭技术，卫星通信由试验阶段进入实用阶段。

1987年8月，中国返回式卫星为法国搭载试验装置。这是中国打入世界航天市场的首次尝试。

1988年9月7日，中国发射一颗试验性气象卫星“风云一号”。这是中国自行研制和发射的第一颗极地轨道气象卫星。

1990年4月7日，中国自行研制的“长征三号”运载火箭在西昌卫星发射中心，把美国制造的“亚洲”1号通信卫星送入预定的轨道，首次取得了为国外用户发射卫星的圆满成功。



1990年7月16日9时40分，中国新研制的大推力运载火箭——“长征”二号捆绑式运载火箭在西昌卫星发射中心发射成功，将模拟卫星送入了预定轨道。这枚火箭是由中国新建的大型航天发射设施发射升空的，同时还为巴基斯坦搭载发射了一颗小型科学实验卫星。

1991年1月22日下午18时23分，中国第1枚120千米高空低纬度探空火箭——“织女三号”在中国科学院海南探空发射场发射试验成功。

1994年2月22日，中国第一座海事卫星地球站通过验收。它的建成填补了中国高科技的一项空白。

1998年5月2日，中国自行研制生产的“长征二号丙”改进型运载火箭在太原卫星发射中心发射成功。这标志着中国具有参与国际中低轨道商业发射市场竞争力。

2003年10月15日，“神舟五号”载人飞船升空；2005年10月12日，“神舟六号”搭载2名航天员升空。2013年6月，“神舟十号”搭载3名航天员升空，其中1名为女性。

2013年4月，诺贝尔奖获得者、华裔科学家丁肇中及其研究团队（包括中国）借助空间阿尔法磁谱仪已发现680万个正电子，正在探索太空的征程中前行。要指出的是，空间阿尔法磁谱仪上的关键部件之一的永磁体是由中国航天人研制的。

5.5 小结

20世纪以来，在前人科学工作的基础上，特别是以牛顿为代表所建立的经典宇宙学引导下，宇宙学获得了巨大的发展。量子力学与相对论的创立，确立了现代宇宙学坚实的基础。宇宙科学理论与技术的进步，催生并促进着航天科技新学科的建立和发展；而航天科技的飞速发展，特别是空间探测技术带来的诸多重大新发现，又进一步给了宇宙理论实验验证和证实的根据。这种良性的相互促动，将使宇宙科学和航天科技继续保持活力，得到更多的新成果。

天体分光技术、照相技术、测光技术和光谱分析技术的进展，直接推动了“赫罗图”的出现，恒星演化研究及其距离的测试、银河系模型的建立、河外星系的确认，极大地拓展了对宇宙的研究；宇宙膨胀学说和哈勃定律（即红移定律）的创立和证实、大爆炸理论的提出、宇宙背景辐射的发现和氦元素丰度的证实，进一步确立了宇宙理论必须要实践来检验和证实的科学方法；宇宙膨胀的终止并转入收缩直至下一次爆炸理论的提出和分析、论证，描述出宇宙未来的图景，期待着科学家们继续提供翔实的观测、计算和仿真的数据，开展更深入的研究。

人造航天器的出现，开创了宇宙研究的新纪元。射电天文望远镜技术、空间天文探测技术、巨型高速计算机及智能技术和大数据信息技术的不断创新，极大地扩展和深化了人类认识宇宙视野，许多前所未有的发现更加充实了人类对真实宇宙的科学认知，同时也直接或间接地造福于人类社会。先进的知识、思想、科学、技术是人类社会不断发展的强大动力和制胜的法宝。可以相信，随着理论工具和技术工具的共同进步，对宇宙的探索会愈加深入，人类能够掌握宇宙的发展规律并提升自

己的素养和能力来适应它，获得更好的长久生存空间。

参考文献

- [1] 钮卫星. 天文学史 [M]. 上海: 上海交通大学出版社, 2011.
- [2] [美] 克里斯琴森. 星云世界的水手——哈勃传 [M]. 何妙福, 朱保如, 译. 上海: 上海科技教育出版社, 2000.
- [3] 小多文化传媒公司. 宇宙深处 [M]. 天津: 天津新蕾出版社, 2013.
- [4] 张邦固. 宇宙中航行 [M]. 北京: 知识产权出版社, 2014.
- [5] 张邦固. 恒星起源动力学 [M]. 北京: 科学出版社, 1994.
- [6] 曹天元. 上帝掷骰子吗? 量子物理史话 [M]. 沈阳: 辽宁教育出版社, 2011.
- [7] [美] 艾萨克森. 爱因斯坦传 [M]. 张卜天, 译. 长沙: 湖南科学技术出版社, 2012.
- [8] [美] 格林. 宇宙的结构 [M]. 刘茗引, 译. 长沙: 湖南科学技术出版社, 2012.
- [9] [英] 考克斯 科恩. 宇宙的奇迹 [M]. 李剑龙, 叶泉志, 译. 北京: 人民邮电出版社, 2014.
- [10] 张邦固. 恒星起源动力学 [M]. 北京: 科学出版社, 1994.
- [11] [英] 帕森斯. 科学的历史 [M]. 涂文文, 译. 北京: 人民邮

电出版社，2014.

[12] [英] 拉森. 霍金传 [M]. 王迪, 译. 上海: 上海远东出版社, 2010.

[13] 苏宜. 天文学新概论 [M]. 北京: 科学出版社, 2012.

[14] 龚钜尔. 航天简史 [M]. 天津: 天津科学技术出版社, 2012.

[15] 唐国东, 华强. 翱翔太空: 中国载人航天之路 [M]. 上海: 上海交通大学出版社, 2012.

[1] 1千秒差距=3261.6光年。

[2] u: 原子质量单位。1u=1.6605×10⁻²⁷kg。